2015

Structure and Molecular Mechanism of a PLP/GABA Dependent Transcription Regulator GabR

Rui Wu
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

STRUCTURE AND MOLECULAR MECHANISM OF A PLP/GABA-DEPENDENT TRANSCRIPTION REGULATOR GABR

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN CHEMISTRY

BY

RUI WU

CHICAGO, IL

MAY 2015

ACKNOWLEDGMENTS

I would like to begin by expressing my sincerest gratitude to my supervisor Dr. Dali Liu, for introducing me and giving me the opportunity to work in the exciting field of X-ray Crystallography. Dr. Lu was a great and patient mentor who provided immense support throughout my Ph.D. studies by introducing me to tons of great collaborators from who I absorbed tons of knowledge, and for everything that he taught me in these years, both in academia and in general.

I am also forever grateful to Dr. Dr. Ruslan Sanishvili at the Argonne National Laboratory and Dr. Miguel A. Ballicora.

Dr. Ruslan Sanishvili provided great support on six or more projects I've being working on throughout these years and his insightful advice made me a better person. His sharp expertise in X-ray crystallography not only has greatly enriched this dissertation, but also taught me how to approach scientific hypothesis with careful and logical steps. On a personal note, I am forever grateful to him for being a listener when I had doubts of myself, and for giving me a confidence boost when I seemed to lose it.

I would also like to thank Dr. Miguel A. Ballicora for his insightful knowledge on enzyme kinetic and showing me how to produce scientific writing in depth. Chapters 5 and 6 could not be accomplished without the great support and tremendous effects from Dr. Ballicora. More importantly, I have benefited a lot from his artistic approach of encouraging and supporting me at critical moments during this long journey.

I would also like to thank Dr. Till Boecking and his graduate student Walid Al-Zyoud at the University of New South Wales in Australia for a lot of help in determining the Small Angle X-ray Scattering Structures of GabR and GabR and DNA complexes. Those structures have shed light on interpreting the data from the X-ray structures and the DNA binding results. It also paves the way for the future functional study on the GabR, DNA complexes with RNA polymerase, sigma factor, etc.

I would also like to thank Dr. Boris R. Belitsky from Tufts Medical School for hosting me at his lab for almost one month and for his great support on the *in vivo* and *in vitro* aspects of the project. I also benefited a lot from his insightful criticisms on my manuscripts.

I would also like to thank Dr. Quyen Quang from Indiana Medical School for hosting my co-workers and me multiple times at his research lab and assisting with the design of multiple experiments and result interpretations. I am also very grateful for many useful life tips and taking me to the conferences with his research team.

I would also like to thank Dr. Richard Holz, Dr. Richard B. Silverman and Dr. Walter Fast for the opportunities to work on their interesting projects, and their constant guidance and suggestions during the process. I also appreciate the helpful advice from Dr. Jacob Ciszek during the job searching process and Dr. Catherine Putonti for serving on my research committee.

I'm forever grateful to Dr. Carlos M. Figueroa and Dr. Salette Martinez Martinez for teaching me valuable skills at the beginning of graduate school, which have tremendously helped me throughout my research. The postdocs: Hoang V. Lee and Jianbin Zheng and graduate students: Yuguang Wang, Zheng Zhang, Ben Hill, Hyun Lee

and Ken Cleavage, thanks for being such good lab mates and collaborators. I'm also grateful to all the undergraduate students who helped me with my experiments: Hansol lee, Michael Farley, Emily Cybulla, and summer intern Saba Sanishvili, and Nushrat Hoque.

I also want to thank the Chemistry and Biochemistry Department main office staff, Carol Grimm, Stacey Lind, and Denise Hall for their help with paperwork matters; and Matthew Sara from the Stockroom for his help with laboratory / chemicals matters.

I would like to acknowledge my family members - my mum and dad, and my friends: Daniel Kechker, Ronan Garcia, Ambrozewski M. Rachel and Zulfiya Gaziev, for years of lessons and amazing support and encouragement, which have enabled me to complete this dissertation and degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF ABBREVIATIONS

GABA	γ-amino butyric acid
wHTH	winged Helix-Turn-Helix
PLP	Pyridoxal 5′-Phosphate
AFPA	(S)-4-Amino-5-Fluoropentanoic Acid
CF	Cystic Fibrosis
Bcc	*Burkholderia cepacia Complex*
B. subtilis	*Bacillus Subtilis*
B. cenocepacia	*Burkholderia Cenocepacia*
B. multivorans.	*Burkholderia Multivorans*
TFs	Transcription Factors
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
PWM	Parallelism Weight Matrix
GntR	Repressor of the Gluconate Operon
HTH	Helix-Turn-Helix
Pfam	Protein families database
E.coli	*Escherichia coli*
DBD	DNA Binding Domain
EB/o	Effector Binding and / or oligomerization domain

PDB	Protein Data Bank
CoA	Coenzyme A
RMSD	Root Mean Square Deviation
ABC	ATP-Binding Cassette
GabT	GABA Aminotransferase
SSDH	Succinic Semialdehyde Dehydrogenase
GabD	Succinic Semialdehyde Dehydrogenase
PMP	Pyridoxamine 5'-Phosphate
NAD	Nicotinamide Adenine Dinucleotide
SS	Succinic Semialdehyde
DNase	Deoxyribonuclease
AT	Aminotransferase
Asp-AT	Aspartate Aminotransferase
bp	Base Pair
NMR	Nuclear Magnetic Resonance
BLAST	Basic Local Alignment Search Tool
FP	Fluorescence Polarization
TAMRA	Tetramethylrhodamine
FP_{FTD}	Fluorescence Polarization Signal of Free Labeled DNA
FP_{GTD}	Fluorescence Polarization Signal of GabR−DNA Complex
K_D	Dissociation Constant
DEG	Diethylene Glycol
UV	UltraViolet

PRT	PhosphoRibosyl Transferase
LB	Lysogeny Broth
IPTG	Isopropyl-β-D-Thiogalactopyranoside
Ni	Nickel
NADP	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Dehydrogenase
SAXS	Small Angle X-ray Scattering
GT-B	Retaining Glycosyltransferase Family B
Suc	Sucrose
UDP	Uridine Diphoshpate
UDP-Glc	Uridine Diphoshpate Glucose
Fru-6P	Fructose-6-phophate
ADP	Adenosine Diphosphate
ADP-Glc	Adenosine Diphosphate Glucose
NDP	Nucleoside Diphosphate
NDP-Glc	Nucleoside Diphosphate Glucose
N. europaea	*Nitrosomonas europaea*
*Neu*SucSase	*Nitrosomonas europaea* SucSase
*Ath*SucSase	*Arabidopsis thaliana* SucSase
CRC	Comprehensive Cancer Center
BSA	Bovine Serum Albumin
NCBI	National Center for Biotechnology Information database
PEG	Polyethylene Glycol

SSN-1	Sucrose Synthase N-terminal-1
SSN-2	Sucrose Synthase N-terminal-2
T. elongatus	*Thermosynechococcus elongatus*
Alanine	Ala/A
Arginine	Arg/R
Asparagine	Asn/N
Aspartic Acid	Asp/D
Cysteine	Cys/C
Glutamic Acid	Glu/E
Glutamine	Gln/Q
Glycine	Gly/G
Histidine	His/H
Isoleucine	Ile/I
Leucine	Leu/L
Lysine	Lys/K
Methionine	Met/M
Phenylalanine	Phe/
Proline	Pro/P
Serine	Ser/S
Threonine	Thr/T
Tryptophan	Trp/W
Tyrosine	Tyr/Y
Valine	Val/V

Adenine	A
Cytosine	C
Guanine	G
Thymine	T

ABSTRACT

GabR is a member of the MocR/GabR subfamily of the GntR family of bacterial transcription regulators. It regulates the metabolism of γ-aminobutyric acid (GABA), an important nitrogen and carbon source for many bacteria. The crystal structures reported here confirmed that this protein has evolved from the fusion of a type I aminotransferase and a winged helix-turn-helix (wHTH) DNA binding protein to form a chimeric protein that adopts a dimeric head-to-tail configuration. The pyridoxal 5′-phosphate (PLP)–binding regulatory domain of GabR is therefore an example of a coenzyme playing a role in transcription regulation rather than in enzymatic catalysis. Our structural and biochemical studies have laid solid the mechanistic foundation for understanding the regulatory functions and mechanism of the MocR/GabR subfamily of transcription regulators.

GabR is also an intriguing case of molecular evolution, displaying the evolutionary lineage between aminotransferase and a resulted regulation domain of a transcription regulator. PLP's native function is not a catalytic co-enzyme but an effector of transcription regulation. The stable chemical species of GabR-PLP-GABA responsible for GabR-mediated transcription activation, has been revealed mechanistic crystallography, NMR spectroscopy and biological assays using both GABA and a

GABA analog (S)-4-Amino-5-Fluoropentanoic Acid (AFPA) as a molecular probe. Our studies provide critical structural and chemical insights for a currently understudied subfamily of bacterial transcription regulators, MocR/GabR-type regulators in the GntR family.

In parallel, I also worked on several other projects. Herein, I dedicated 2 full chapters to demonstrate the progress on the Sucrose Synthase project. We obtained biochemical and structural evidence of sucrose metabolism in non-photosynthetic bacteria. Until now, only sucrose synthases from photosynthetic organisms have been characterized. Here, we provide the crystal structure of the sucrose synthase from the chemolithoautotroph *Nitrosomonas europaea*. This supported that the enzyme functions with an open/close induced fit mechanism. It prefers as substrate adenine-based nucleotides rather than uredine-based like the eukaryotic counterparts, implying there is a strong connection between sucrose and glycogen metabolism in these bacteria. Mutagenesis data showed that the catalytic mechanism must be conserved not only in sucrose synthases, but also in all other retaining GT-B glycosyltransferases.

CHAPTER ONE

INTRODUCTION

Research Aims: Potential Anti-Virulence Target for Secondary Infections of Cystic Fibrosis (CF)

Cystic Fibrosis is a life-threatening genetic disorder that causes thick, sticky mucus to build up in the lungs, digestive tract, and other areas of the body. It is one of the most common chronic lung diseases among children and young adults. In Caucasian populations, the carrier frequency for Cystic Fibrosis is approximately one in twenty-five, with a prevalence of about 1 in 2000 [1]. Nearly 1,000 CF diagnoses are reported annually, and the disease caused 3,708 deaths in the United States from 1999-2006. As of 2007, the medial survival age was 37.4 years, compared to a U.S. life expectancy of 70.7 years in the same year.

The local environment of the cystic fibrosis lung is unique and results in characteristic disease patterns and specific spectrum of cystic fibrosis associated lung pathogens [1]. Susceptibility to *Staphylococcus aureus*, *Pseudomonas aeruginosa* and other bacterial infections is significantly higher in patients with CF, as the lungs become largely ineffective at ridding the body of microbes when mucus build-up begins. Thus, a large number of deaths from CF are attributed to bacterial secondary infections that lead to lung disease in patients. With CF affecting 70,000 people across the world, and the bacteria's high resistance of a wide range of antibacterial agents, the novel antibiotic

development becomes urgent for the treatment of bacterial infections and for an improved life quality of CF patients.

The *Burkholderia cepacia Complex (Bcc)* is a subset of at least nine opportunistic pathogens that belong to the gram-negative *Burkholderia* genus associated with infections in CF patients. In efforts to develop more effective CF treatments, this complex is of particular interest in antibiotic development because of the well-documented multi-drug resistance of bacterial strains and the capability of patient-to-patient spread. *Burkholderia cenocepacia* and *Burkholderia multivorans* are two members of the *Bcc* responsible for a majority of the infections seen in CF patients, and they have been reported in both clinical and environmental settings (soil and water) as sources of infection.

These two strains are related to *Bacillus subtilis* by their metabolism of γ-amino butyric acid (GABA) to produce glutamate, which is an important metabolite in bacterial nitrogen metabolism. Similarity existing between *B. subtilis* and Bcc strains in their glutamate production mechanisms makes *B. subtilis* an ideal, non-pathogenic model organism for our study on the inhibition/modulation of GABA/glutamate metabolism in *B. cenocepacia* and *B. multivorans* for therapeutically purposes. It also has been recently discovered that GABA is an essential cell-to-cell signal in eukaryotes. It acts as a single signal between eukaryotes and pathogenic bacterial. GABA production has been revealed to be crucial to bacterial stress response, especially acid resistance. All of which make the metabolism of GABA in bacteria a potential target for the development of a strategy to control the virulence of bacterial pathogeneses. GabR is the transcription regulator in *B.*

subtillis that represses the transcription of the gabR gene while also initiating the transcription of the gabT and gabD genes, which are key enzymes in glutamate biosynthesis utilizing GABA while responding to the increase of GABA's concentration *in vivo.*

The Transcription and the Transcription Factor

As the first step of gene expression, transcription is a fundamental process that is often highly conserved and tightly regulated process in bacteria. Bacteria cells rely on very sophisticated responses to adapt to micro environmental changes, due to the frequently changing habitats conditions. Though transcription regulation, these responses most frequently lead to the activation and/or repression of a specific sets of genes to control metabolic pathway and adapt to the different environmental condition [2].

Transcription is the process of synthesis of mRNA by using a single strand of DNA as a template; the process is performed by structurally conserved DNA-dependent RNA polymerase (RNAPs) [3]. Transcription is tightly controlled by the transcription factors (TFs), which are the proteins that bind to specific sequence on the DNA near their target genes, thereby modulating the transcription initiation. TFs perform this function alone or with other proteins in a complex (homoprotein), by promoting (as an activator), or blocking (as a repressor) the recruitment of RNA polymerase to the promoter regions of specific genes.

Based on contemporary X-ray crystallography and advanced computational simulation, TFs mostly revealed as the dimerization motif, both heterodimer and homo-dimer are common. In general, the dimerization of a protein generally results in the

formation of an enlarged interaction surface, compared to the monomer. The enlarged surface area of the dimer provides increased potential for various protein-protein or protein-DNA interactions essential to transcription. Comparing to a monomer, TFs that dimerize can achieve a higher DNA binding affinity and specificity and twice as many base pairs can be recognized providing distinct DNA-binding specificities [4].

Taking advantage of the intense and ongoing computationally studies, molecular and biophysical studies over the past decade, scientists have generated an unprecedented amount of information describing the location of the binding site/s, described in Parallelism Weight Matrix (PWM) and further deduced from aligning a set of DNA sequence that are experimentally known, for the transcription regulators [5]. However the understanding of the binding mechanism is still very preliminary. The major challenge for our work is to be able to bridge this gap between the molecular description and the genome-scale overview.

GntR family and the subfamily classification

In 1990, Haydon and Guest first identified and described GntR family of bacterial transcription regulators, named after the first member identified, the *B. subtilis* transcription regulator - GntR - repressor of the gluconate operon [6]. Among all the bacterial transcription regulators, GntR is known as one of the most abundant and widely distributed groups of Helix-Turn-Helix (HTH) transcription factors. To date, the metabolite-response GntR family has approximately 8,561 regulatory proteins in Pfam database (PfaM family GntR: PF00392)[7], which spread out among 764 bacteria taxa. GntR transcription regulators are often located at the chromosome adjacent to the genes

that they control, which in many cases allows insight into the metabolites that they bind to as effectors. There are also many cases where GntR-like regulators are not located close to the gene it regulates but elsewhere. One of the well-studied examples in this transcription regulator family is orphan regulator FadR/GntR, the fatty acid metabolism regulator in *Escherichia coli*. These studies have intensively focused on both the structural and functional aspects, serving as a critical reference case for our research.

A striking characteristic feature of the GntR family is that the members possess a highly conserved N-terminal DNA-binding domain (DBD), which evolved in DNA binding, and an extensive diverse and heterogeneity C-terminal effector binding and / or oligomerization domain (Eb/O). It was hypothesized that binding of the effector molecules could trigger the conformational changes in the protein, which modulates the DNA-binding properties of the regulator resulting in either activation or repression of transcription. It is not surprising, though, that there are a large number of effector molecules that have not been identified, given the diversity of molecule that they bind and relative large number of the GunR-family members itself.

The HTH DNA binding motif in this GntR family is the traditional tri-helical domain, whose original case demonstrated the critical role in DNA binding within the bacteirophage *Lambda* protein, *cI* and *cro* and the *lac* operon repressor, *LacI* and the generation of cell-type specific gene expression [8]. The there-helical HTH domain is known as the winged-helix-turn-helix domain. In addition to the core three-helix bundle, it also contains a wing form by C-terminal beta-hairpin. In some versions of this domain, the loop between strand elements interacts with the two strands of the wing [9]. The “wing”

feature provides great interactions with the DNA due to the third alpha helix, which can fit well within the major groove of the DNA. This was suggested by the extensive sequence analysis and secondary structural analysis throughout 1980s and 1990s [7]. It is being proposed that these domains descend from a common ancestor.

Since the physiological role of these regulators critically depend upon the effector binding to the operon site of the DNA, therefore analysis based on the alignment of the C terminus and secondary structural predications were used to classify the members of the GntR family. The distinct structural topologies reveal seven major subfamilies: FadR, HutC, MocR, YtrA, and three minor subfamily AraR, DevA, and PlmA (Figure1) [10].

The most abundant GntR sub-family, FadR subfamily, named after the fatty acid biosynthesis and degradation regulator, functions as a switch that coordinately regulates the machinery required for fatty acid beta-oxidation and the expression of a key enzyme in fatty acid biosynthesis. It contains the majority of the GntR-like regulators, approximately 40 %. The protein in this subfamily contains six or seven-helical C-terminal domain, with the average length of 150-170 amino acid. The crystal structure is available for FadR of *Escherichia coli* (Figure2; PDB ID: 1EX1)[11].

The HTH domain is fused to a seven alpha helical bundle, which contains a large internal cavity required for binding its regulator molecule. The RMSD on the alpha carbons is 0.717 Å between the bound and unbound of the acyl CoA indicated the protein backbone undergoes conformational shift upon binding to the acyl CoA-its own regulator, further resulting in 7.2 Å movement of the DNA recognition helix preventing DNA binding and subsequently repressing the transcription.

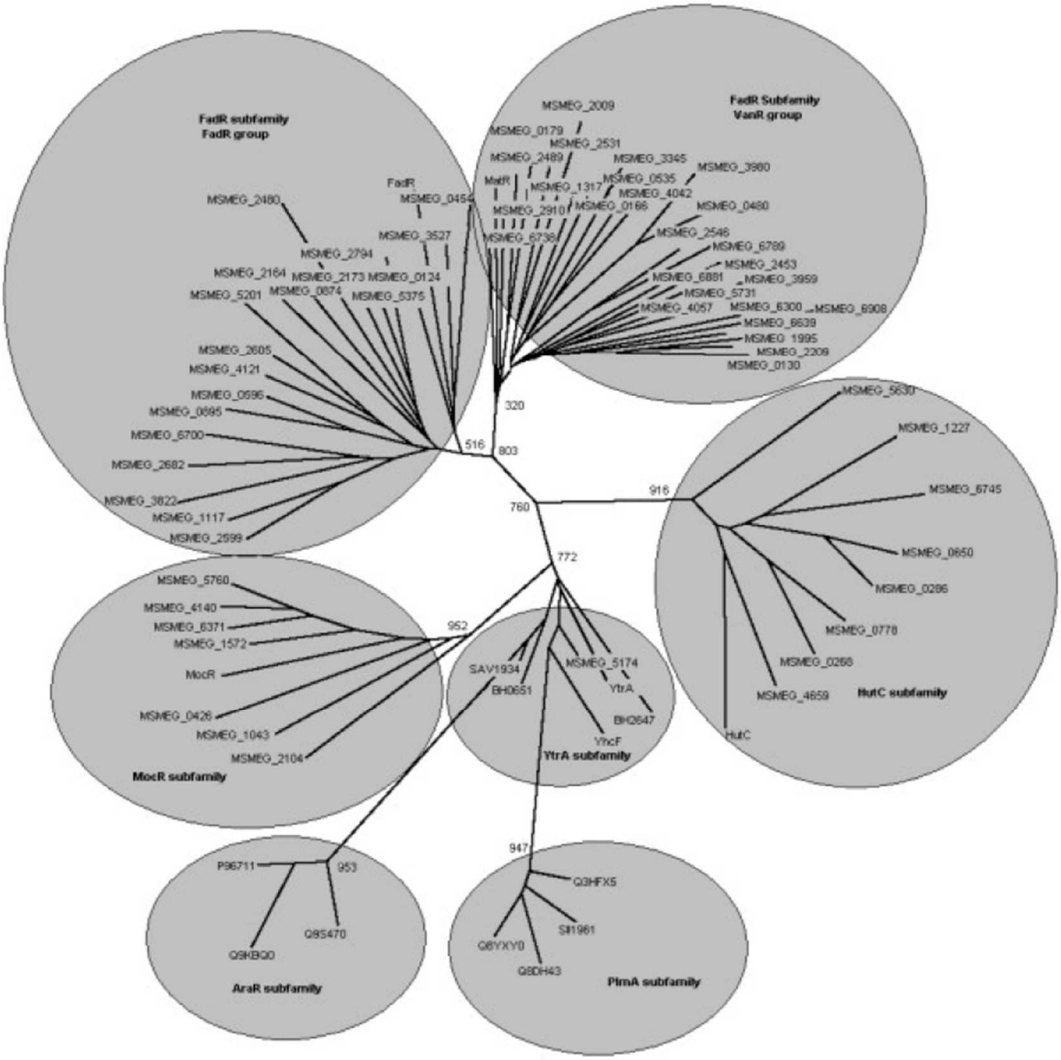


Figure 1. Unrooted tree of the proteins of GntR family regulators of *M. smegmatis* including representatives of all subfamily regulator from different Baterial Genomes with 1000 bootstrap replicates. All the GntR regulators are clustered into six subfamilies. FadR subfamily is branched again into two groups (FadR and VanR) [10].

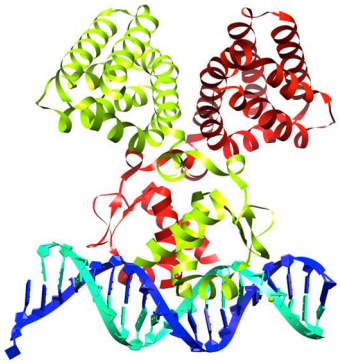


Figure 2. Quaternary structure of FadR and its corresponding DNA.

The second proposed subfamily is the HutC subfamily, named after the histidine utilization operon regulator, comprising about 31 % of the GntR regulators. The C-terminal regulatory domain of HutC response to various of the ligands, such as: histidine [12]; long chain fatty acids [13]; trehaylose 6-phosphate or alkylphosphonate [14], which makes it not so surprising to find out that this subfamily regulate processes as diverse as amino acid uptake and plasmid transfer. The structure of HutC subfamily contains both alpha-helical and beta-sheet on C-terminal domain and averages 170 amino acids in length; Three dimensional crystal structural and analysis of the HutC family member, PhnF from E. coli shows that the C-terminal domain had a chorismate lyase-like fold, comprising a six-stranded antiparallel beta-sheets, a two-stranded parallel beta-sheets and four short alpha-helices (Figure3; PDB ID: 2WV0). A large cavity on the surface of the PhnF C terminus indicates the putative effector molecule-binding pocket, which suggested to be conserved in with the other members in the HutC subfamily [15].

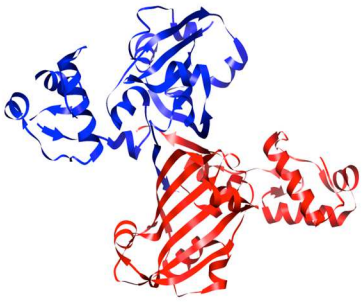
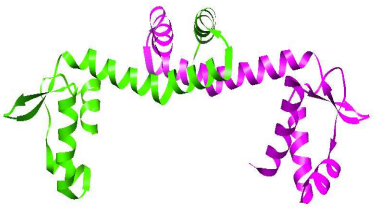


Figure 3. Quaternary structure of PhnF.

The YtrA sub-family represents only 6 % GntR-like regulators to date, among which forms part of operons involved in ATP-binding cassette (ABC) transport system. This sub-family contains the shortest C-terminal domain with about 50 amino acids on average (Figure 4), increasing certain extents of doubts regarding the efficiency for binding of the effecter molecule(/s). People hypothesizes that the binding of the small molecules may impair the dimerization of this particular family of the transcription regulators, which could cause the loss of DNA binding, thus further repress the transcription. Combining with the sequence comparison and the secondary structural predication, it was suggested that dimerization is quite necessary for this subfamily and



could be formed by positively

Figure 4. Quaternary structure of CFL2947 from *Corynebacterium glutamicum*.

or negatively charged as well as hydrophobic and aromatic residues[16]. This hypothesis soon proved by the X-ray structure of CFL2947 from *Corynebacterium glutamicum* (Figure 4; PDB ID: 2EK5)[17].

The DevA sub family of GntR transcription regulators has only be found in bacterium *Streptomyces* [18]. It showed that DevA is required for proper developments in *S. coelicolor,* although the exact roles of these proteins have not been highly elucidated. The C-terminal domain topology is clearly novel comparing to the others subfamilies. The N-terminal domain shares the most similarity with the HutC sub family [18]. Since there are no crystal structures available in this subfamily, the structure-function relationship remains unclear.

AraR is another minor subfamily, which required for the extracellular degradation of arabinose-containing polysaccharides, transport of arabinose, arabinose oligomers, xylose, and galactose, intracellular degradation of arabinose oligomers, and further catabolism of this sugar in *B. subtilis* [19]. The C-terminal domain shows extensive homology to the LacI/ GaIR regulator family. The crystal structure of the AraR (Figure 5; PDB ID: 4EGY) demonstrated a high-level repression by co-ordinarily binding two-in-phase [20], which might correlate with different physiological requirements. The high level of repression is achieved by DNA bending requiring two in-phrase, whereas binding to a single operator, which autoregulates AraR expression, is 10-fold less effective [21].

Figure 5. Quaternary structure of DNA binding domain from AraR and its corresponding DNA fragment.

PlmA subfamiy, the most newly discovered minor subfamily in 2003 that found in cyanobacterial *Anabaena sp.* Strain PCC7120. It has been demonstrated that plmA can affect on a cell's ability to maintain its relative plasmid content. However, regulation mechanism of PlmA remains unclear [22]. Structurae wise, PlmA also has a novel topology on the C-terminal domain, and appears to confine to the Cyanobacateria. The N-terminal DNA binding domain appear to be closely similar to the MocR and YtrA subfamilies suggesting that the E-b/O domain of PlmA may have been divergently replaced during the evolution, despite their similarity in the DNA binding domain.

The MocR subfamily's C-terminal domain is readily distinguishable from others because of its exceptional average length of about 350 amino acids. It is also capable of binding to a vitamin B6 derivitive-PLP. The PLP-binding putative aminotransferase domains are closely related by amino acid sequence to the homodimeric type-I aminotransferases. Aminotransferases are of the central importance in amino acid metabolism and in links to carbohydrate and fat metabolism. Type-I aminotransferases almost excusively exist as dimers in a head-to-tail fashion, and each subunit binds to one molecule of PLP through an internal aldimine linkage with the amino group of the

conserved lysine residue in the PLP binding site. This characteristic feature suggests that the recruitment of such a domain in transcription regulation is likely to facilitate protein-protein interaction/oligermerization.

MocR/GabR

The representative member of the MocR/GabR subfamily, GabR, is essential for transcriptional activation of the pathway that controls the GABA degradation [23]. GABA is an important nitrogen and carbon source in many bacteria. The GABA-inducible *gabTD* pathway in *B. subtilis* allows utilization of extracellular GABA as the sole nitrogen source to generate glutamate and succinate. The *gabR* gene in *B. subtilis* is located upstream of the *gabTD* operon and is transcribed divergently. GabR functions as the PLP- and GABA-dependent transcriptional activator of the *gabTD* operon and as a negative autoregulator [3](Figure 6). Mutations in the conserved residues

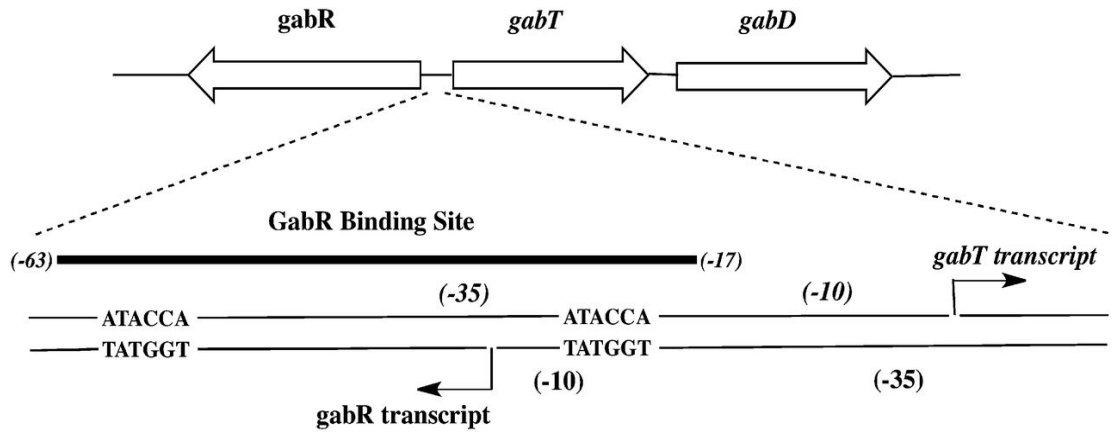


Figure 6. Schematic of GabR binding sites at the *gabR* and *gabTD* promoters. A schematic of the *gabR* and *gabTD* gene loci, as previously elucidated (2), is shown here. The 47-bp GabR binding site (−63 to −17) within the *gabR*−*gabT* regulatory region is shown in detail, and the two direct repeats (ATACCA) are indicated. The arrows indicate the transcription start sites and directions for the *gabR* and *gabT* genes

of the putative aminotransferase domain presumed to be crucial for the folding of the domain and PLP-binding abolished the ability of GabR to activate the *gabT* promoter in vivo but did not affect its activity as an autorepressor [23]. Therefore, while the ability to bind DNA may be sufficient for its autorepressor function, GabR is competent to activate transcription from the *gabT* promoter only in the presence of PLP and GABA.

The *gabT* and *gabD* genes encode the two key enzymes for glutamate biosynthesis: GABA aminotransferase (GabT) and succinic semialdehyde dehydrogenase (SSDH or GabD), respectively. GabT is a type-I aminotransferase that catalyzes the Ping-Pong transamination reaction using GABA as amino group donor, α-ketoglutarate as amino group acceptor and PLP as co-enzyme to form the products, succinic semialdehyde and glutamate [24]. During the first half-reaction, GABA donates its amino group to PLP, thereby releasing succinic semialdehyde and converting PLP to its pyridoxamine 5'-phosphate (PMP) form (Figure 7). In the second half-reaction, α-ketoglutarate accepts the amino group to become glutamate and regenerate the PLP form of the enzyme (Figure 7). In the downstream reaction, the NAD-dependent dehydrogenase, GabD converts succinic semialdehyde (SS) to succinate. Null mutants of gabR, gabT and gabD in *B. subtilis* were reported to be incapable of using GABA as the sole nitrogen source [25], further demonstrating that the functions of GabR, GabT and GabD are all essential for GABA utilization.

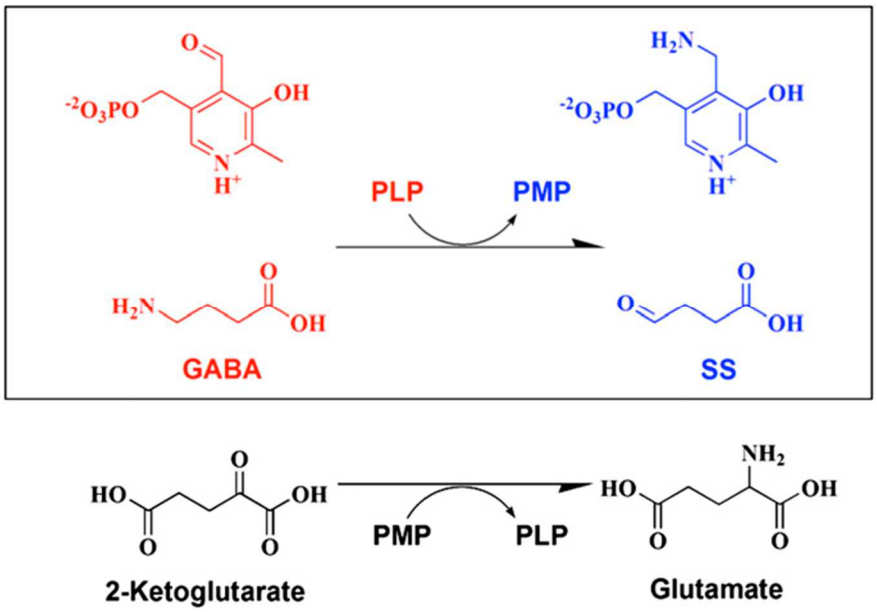


Figure 7. The GabT-catalyzed aminotransferase reaction. The first half reaction (GABA to succinic semialdehyde) of the GabT-catalyzed Ping-Pong transamination is shown in the boxed area. The second half-reaction is shown outside the boxed area.

CHAPTER TWO

CRYSTAL STRUCTURE OF BACILLUS SUBTILIS GABR, AN AUTOREPRESSOR AND TRANSCRIPTIONAL ACTIVATOR OF GABT

Research Significance

B. subtilis GabR is a transcription factor that regulates gamma-aminobutyric acid (GABA) metabolism. GabR is a member of the understudied MocR/GabR subfamily of the GntR family of transcription regulators. A typical MocR/GabR-type regulator is a chimeric protein containing a short N-terminal winged-helix-turn-helix DNA-binding domain and a long C-terminal pyridoxal 5′-phosphate (PLP)-binding putative aminotransferase domain. In the presence of PLP and GABA, GabR activates the gabTD operon, which allows the bacterium to use GABA as nitrogen and carbon sources. GabR binds to its own promoter and represses gabR transcription in the absence of GABA. Here, we report two crystal structures of full length GabR from B. subtilis: a 2.7-Å structure of GabR with PLP bound and the 2.55-Å apo structure of GabR without PLP. The quaternary structure of GabR is a head-to-tail domain-swap homodimer. As predicted base on sequence similarity, each monomer comprises two domains: an N-terminal winged-helix DNA-binding domain and a C-terminal PLP-binding type I aminotransferase-like domain. The winged-helix domain contains putative DNA-binding residues conserved in other GntR-type regulators. Together with sedimentation velocity and fluorescence polarization assays, the crystal structure of GabR provides insights into

DNA binding by GabR at the gabR and gabT promoters. The absence of GabR-mediated aminotransferase activity in the presence of GABA and PLP, and the presence of an active site configuration that is incompatible with stabilization of the GABA external aldimine suggest that a GabR aminotransferase-like activity involving GABA and PLP is not essential to its primary function as a transcription regulator.

Introduction

The MocR/GabR subfamily of chimeric proteins is widespread in bacteria and regulates a variety of biological processes. All members of this subfamily are typically composed of a 60- to 120-residue N-terminal winged-helix DNA-binding domain resembling other bacterial transcriptional regulators from the GntR family, and a large C-terminal putative aminotransferase domain, averaging □300 aa in size and capable of binding to the vitamin B6 coenzyme, pyridoxal 5′-phosphate (PLP) [25]. The PLP-binding putative aminotransferase domains are closely related by amino acid sequence to homodimeric type I aminotransferases [26,27]. GABA is an important nitrogen and carbon source in many bacteria. The GABA-inducible gabTD pathway in Bacillus subtilis allows utilization of extracellular GABA as nitrogen and carbon sources. The gabR gene is located upstream of the gabTD operon and is transcribed divergently. GabR functions as the PLP- and GABA-dependent transcriptional activator of the gabTD operon and as a GABA-independent negative autoregulator (figure 6). [24, 25]. Mutations in the conserved residues of the putative aminotransferase domain presumed to be crucial for the folding of the domain and PLP-binding abolished the ability of GabR to activate the gabT

promoter in vivo but did not affect its activity as an autorepressor [25].

The gabT and gabD genes encode two enzymes for an alternative route of glutamate biosynthesis using GABA: GABA aminotransferase (GabT) and succinic semialdehyde dehydrogenase (GabD), respectively [25]. GabT is a type I aminotransferase [28, 29] that catalyzes the Ping-Pong transamination reaction using GABA as an amino group donor, α-ketoglutarate as an amino group acceptor and PLP as coenzyme to form the products, succinic semialdehyde (SS) and glutamate. During the first half-reaction,

GABA donates its amino group to PLP, thereby releasing SS and converting PLP to pyridoxamine 5′-phosphate (PMP) (Figure 7). In the second half-reaction, α-ketoglutarate accepts the amino group became glutamate and regenerates the PLP form of the enzyme (Figure 7). In the downstream reaction, the NAD-dependent dehydrogenase, GabD converts SS to succinate. Null mutants of gabR, gabT, and gabD in *B. subtilis* were reported to be incapable of using GABA as the sole nitrogen source [25].

GabR binds specifically to an extended region of DNA that overlaps the −35 and −10 elements of the gabT and gabR promoters, respectively [24](Figure 7). Using DNase I foot printing experiments and native gel-shift experiments with purified GabR, a 47-bp region of DNA (−63 to −17 relative to the gabT transcription start site) was identified as the GabR binding site. The 47-bp minimal fragment contains two copies of the direct repeat, ATACCA, which may be essential for GabR binding. Addition of PLP and GABA did not noticeably alter the DNase I footprints, indicating that GabR is capable of binding DNA in the absence of these molecules. Based on the extended 47-bp footprint, it

was proposed that GabR binds to DNA either as a dimer or tetramer [24]. It has been reported that a GabR-catalyzed "partial" aminotransferase- like reaction involving GABA and PLP may be essential for its transcriptional activator function [24]. Interestingly, no "full" two-step Ping-Pong–type GABA aminotransferase activity for GabR could be detected in vivo, and overexpression of GabR was unable to restore GABA aminotransferase activity in strains lacking gabT [24]. Here, we report two crystal structures of the PLP-bound and PLP-free forms of full-length *B. subtilis* GabR determined to be 2.7 and 2.55 Å, respectively. The crystal structure of GabR reveals a two-domain protein organized as a distinct head-to-tail domain-swap homodimer. A long linker of 29 aa separates the N-terminal winged-helix domain from the C-terminal domain, which has the same overall fold as type I aminotransferases and is hereby called AT-fold domain. Together with sedimentation velocity and fluorescence polarization data, the GabR structure provides a model for DNA binding at the gabR-gabT locus by one or more GabR homodimers. Key amino acid sequence changes in the putative substrate-binding pocket of GabR relative to those of well-characterized type I aminotransferases suggest that a GabR catalytic activity involving GABA and PLP is either absent or very poor under physiological conditions. Our results strongly suggest that the AT-fold domain in *B. subtilis* GabR, in a unique combination with the winged-helix domain, is primarily relegated to a regulatory function at the gabR and gabT promoters.

GabR is a Head-to-Tail Domain-Swap Homodimer

Sequence alignments of members of the MocR/GabR subfamily show that these proteins comprise a small N-terminal winged-helix domain and a large C-terminal putative type I aminotransferase domain [16, 25]. The crystal structure of GabR described here is

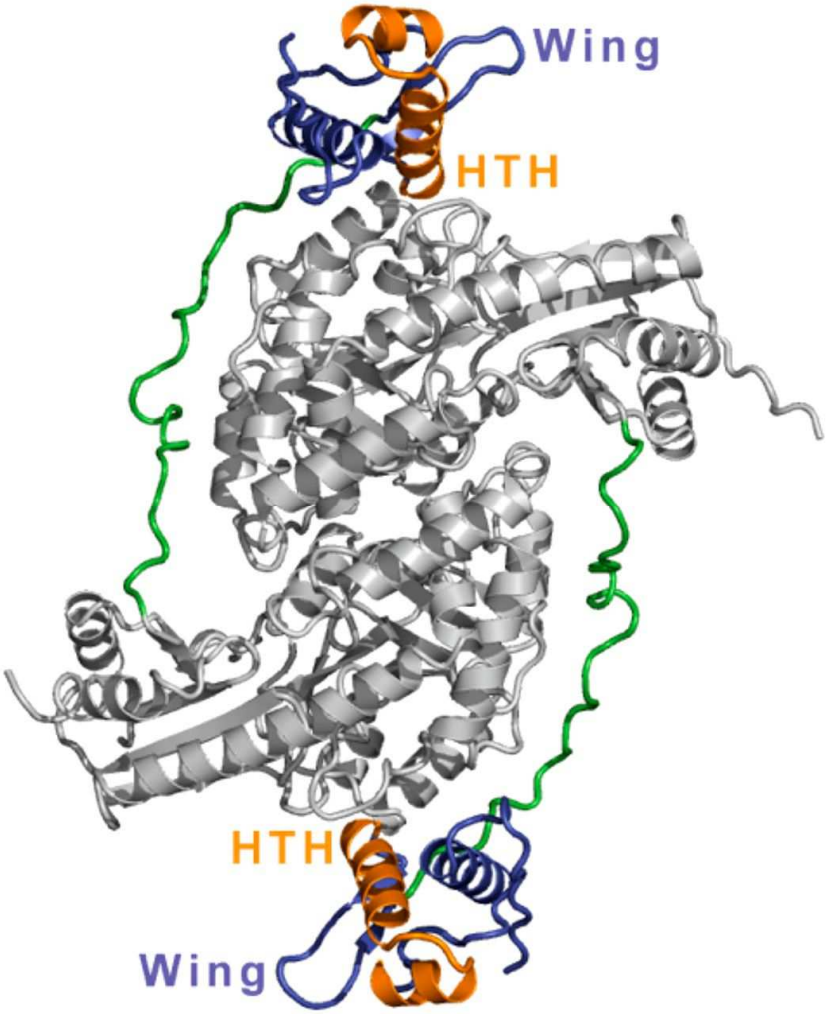


Figure 8. Quaternary structure of GabR. GabR is a head-to-tail domain-swap homodimer. The AT-fold domains are shown in light gray. The long linkers connecting the N-terminal winged-helix domains to the C-terminal AT-fold domains are shown in green. The

winged-helix domains are shown in blue-violet. The HTH motifs (orange) and the wing motifs are labeled. An orthogonal view of the structure is shown in Figure 10.

the winged-helix subtype [30] and a C-terminal AT-fold domain (residues 110–470) characteristic of the type I aminotransferase family [31, 32] and containing the signature long bent α-helix (residues 345–386) connecting two subdomains. A long 29-residue linker (residues 81–109) connects the winged-helix domain to the AT-fold domain (Figure 8 and Figure 9).

Figure 9. Quaternary structure of GabR. The structure is rotated 90° counterclockwise along the y axis relative to Figure 8.

GabR crystallizes as a homodimer with the two monomers organized in the head-to-tail domain-swapped interfaces. The larger interface is composed of two AT-fold domains, which are also arranged in a head-to-tail manner to form the characteristic homodimeric type I aminotransferase fold. The interface contains a buried area of 1,695 Å2 with a binding energy of −11.4 kcal/mol, suggesting that this interface stabilizes the

GabR homodimer and is physiological. This interface of GabR, as seen in all type I aminotransferases [31], contains two PLP-binding pockets bound by one molecule of PLP in each pocket.

The second interface is formed between the winged-helix domain of one monomer and the AT-fold domain of the other monomer. This interface has a buried area of 422 Å2 with a relatively low free energy of −3.8 kcal/mol, suggesting that the overall quaternary structure of GabR may be dynamic in solution and that the winged-helix domain may undergo large-scale domain movements, especially when bound by DNA. The large dimer interface between the two AT-domains further suggests that any such domain movements by the winged-helix domain do not require restructuring of the AT−AT interface. Interestingly, we note that the extended N-terminal arms of the two monomers of aspartate aminotransferase (Asp-AT) also reach over and interact with the other subunit in that homodimer [32]. Because Asp-AT does not have an N-terminal winged-helix domain, the significance of this "crossing over" of the N-terminal arm to the other subunit in Asp-AT is not clear. Although there is reasonable electron density for the long extended linker that connects the winged-helix domain to the AT-fold domain, the linker does not form any significant contacts with the rest of the protein, suggesting that it may be flexible in solution and may thereby facilitate the proposed winged-helixed domain movements upon DNA binding.

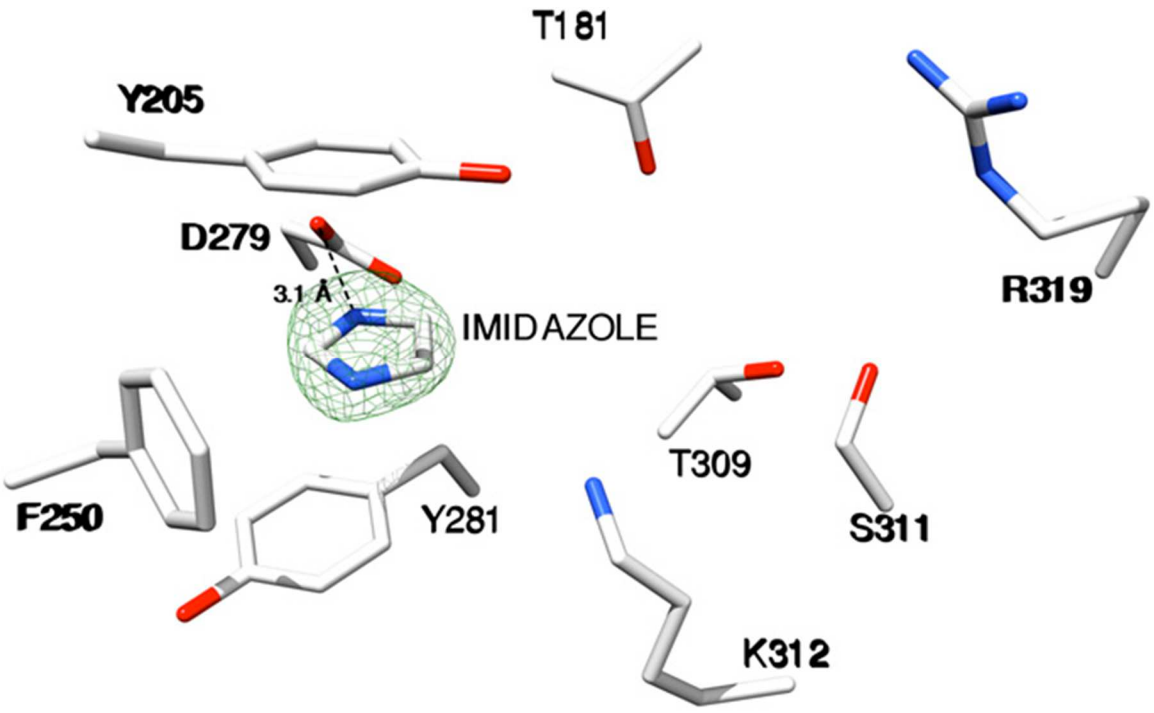


Figure 10. Imidazole is bound in the PLP-binding pocket of apo-GabR. The PLP-binding site of the crystal structure of apo-GabR is shown. The 2Fo − Fc map (green mesh) around the imidazole molecule is shown at 1σ level. The imidazole moiety and surrounding side chains are shown as sticks. Carbon atoms are shown in white, oxygen atoms in red, and nitrogen atoms in blue.

We also crystallized GabR without exogenous PLP. In the resulting apo-GabR structure, PLP is no longer bound in the PLP-binding pocket, and instead, a molecule of imidazole is bound in the same pocket (Figure 10). The purified protein is pale yellow in color, suggesting that endogenous PLP from Escherichia coli is still bound to GabR immediately after purification. Because imidazole is present in high concentrations in the purification and crystallization buffers, the replacement of PLP by imidazole is most likely a crystallization artifact. The overall similarity between the GabR−imidazole and

GabR−PLP structures suggests that the AT-fold domain maintains its overall structural integrity even when PLP is no longer bound.

GabR is a Dimer in Solution

We performed sedimentation velocity analysis to investigate the oligomeric state of GabR−PLP in solution. In the absence of GABA, the sedimentation velocity plot shows a predominant peak at □100 kDa, suggesting a dimer of GabR in solution even at very low protein concentration (Figure 11; GabR molecular weight, 56 kDa). The plot also shows two minor peaks at about 50 and 10 kDa, which may represent a small proportion of monomer and degradation products, respectively. The sedimentation velocity plot of GabR−PLP in the presence of GABA similarly shows a predominant peak at □100 kDa and two low–molecular-weight peaks (Figure 11), again indicating a dimer of GabR in solution. Also, during protein purification, GabR elutes as one major peak on a Superdex 200 gel-filtration column with an apparent molecular weight of a dimer (Figure 12). Based on the crystal structure, our expectation is that the AT−AT dimer interface in the solution state of GabR may be similar to that found in the crystal structure. Our sedimentation velocity data do not reveal tetramers and higher oligomers in solution; however, further oligomerization of GabR may still occur when GabR binds to DNA.

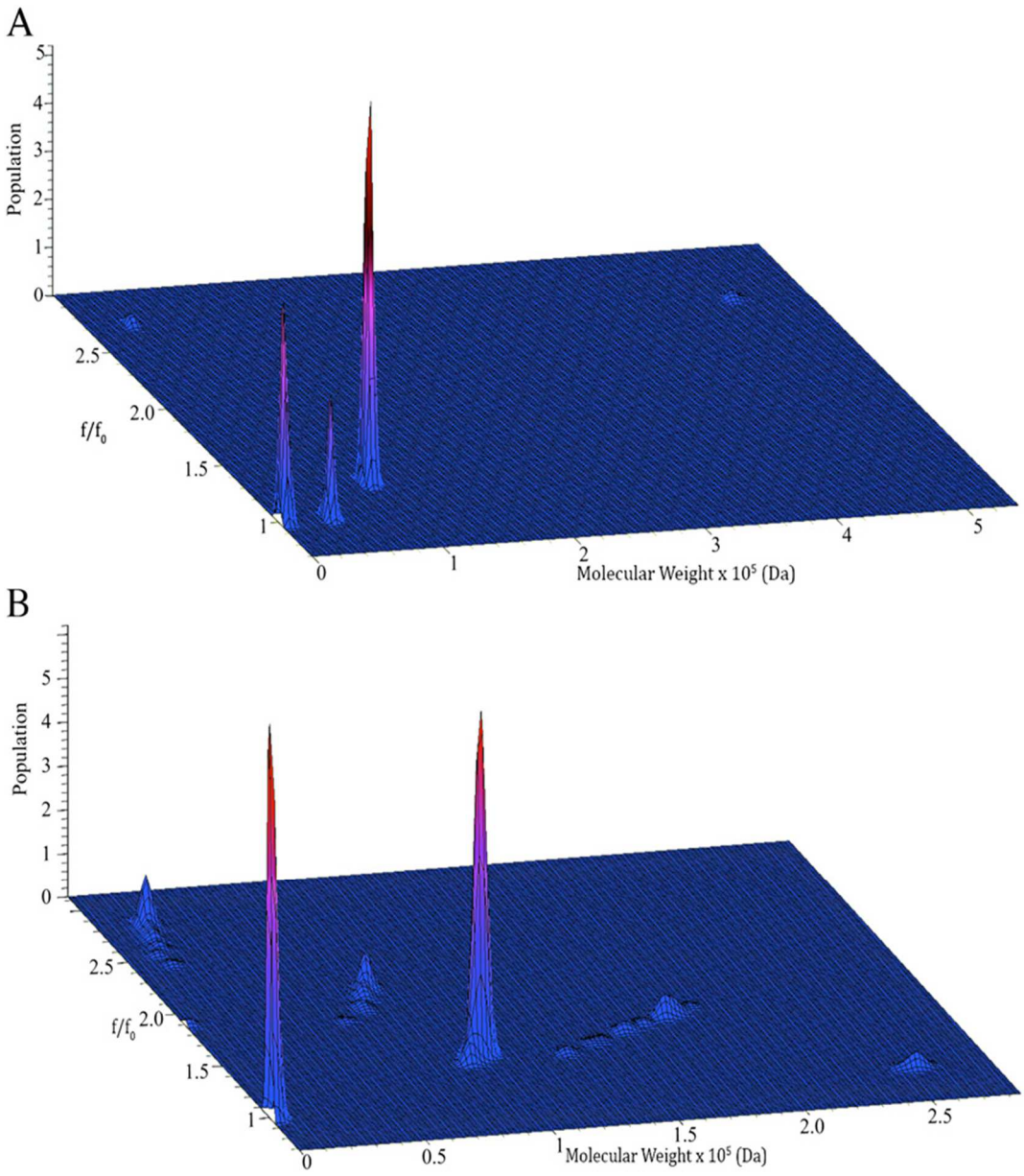


Figure 11. Sedimentation velocity analysis of GabR. (A) Molecular weight of the GabR−PLP sample determined from the global fitting of sedimentation velocity data. (B) Molecular weight of the GabR−PLP sample in the presence of 20 mM GABA

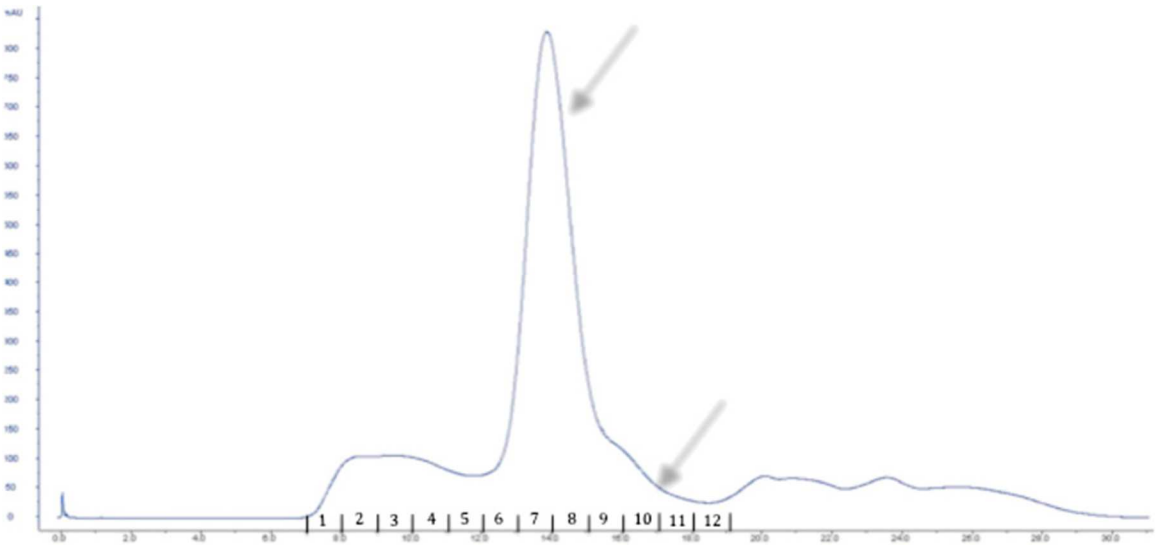


Figure 12. The gel-filtration profile of GabR The major peak for GabR (fractions 7-9) on Superdex 200 16/60 corresponds to an apparent molecular weight of a dimer. The minor peak (fractions 9-10) corresponds to that for a monomer.

Conserved Putative DNA-Binding Residues in the N-Terminal Winged Helix Domain

The GabR N-terminal domain (residues 3–80) contains a canonical winged-helix domain (α1-β1-α2-T-α3-β2-W- β3), where α2-T-α3 (residues 42–66) form the HTH motif and W (residues 71–76) is the wing (Figure 13). In the case of GabR, a second wing is absent. Although many winged-helix proteins contain a second wing, there are several examples of wingedhelix proteins that bind DNA and in which the second wing is absent [33]. X-ray and NMR structures of winged-helix proteins in complex with DNA show that the winged-helix domain uses diverse modes of DNA binding, with examples of this domain binding DNA as a monomer, homodimer, or heterodimer [33].

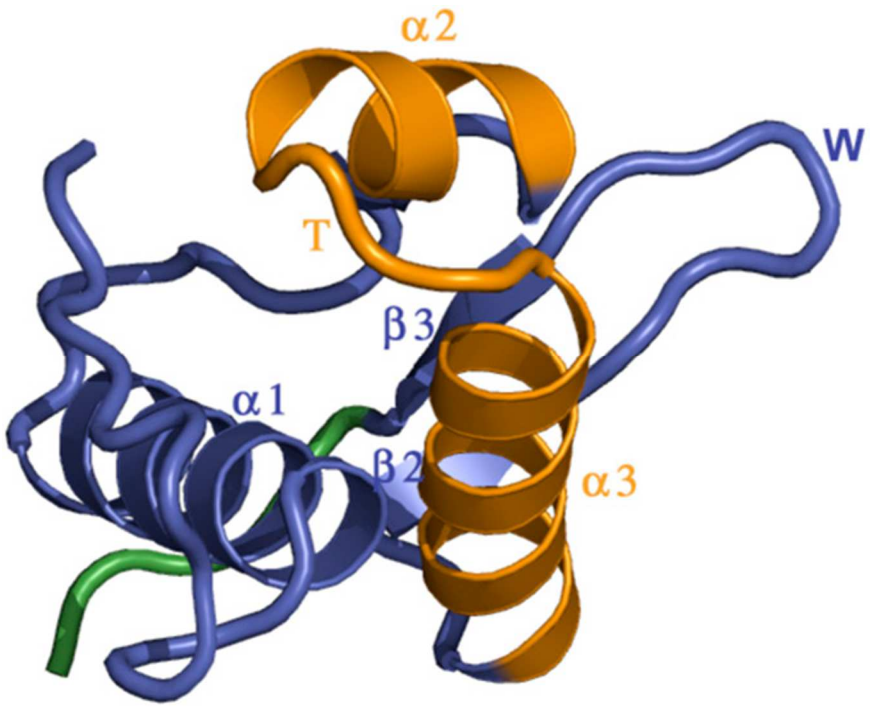


Figure 13. The GabR N-terminal winged-helix domain. TheThe GabR N-terminal winged-helix domain (residues 3–80) is shown in detail. The various secondary structural elements (α1-β1-α2-T-α3-β2-W-β3) are labeled. The helix-turn-helix (HTH) motif (α2-T-α3) is shown in orange. W is the wing (β2-W-β3).

In general, the most basic or positively charged surface of the winged-helix domain binds to the negatively charged phosphate backbone or nucleotide side-chain groups [33]. Because the putative DNA-binding domain is conserved among GntR-type transcriptional regulators, we searched the Conserved Domain Database [34] using the GabR winged-helix domain sequence as a query. We found several hits, including E. coli acyl-CoA–responsive transcription factor, FadR, and identified three putative DNA-binding residues in GabR (R43, S52, and K75) based on sequence conservation in FadR and other GntR family members. Because these three residues (R43 and S52 in the HTH;

K75 in the wing) also map to the most basic surface on the GabR winged-helix domain (Figure 14) and are fully exposed in the GabR homodimer (Figure 15), we predict that they may constitute a major DNA-binding surface.

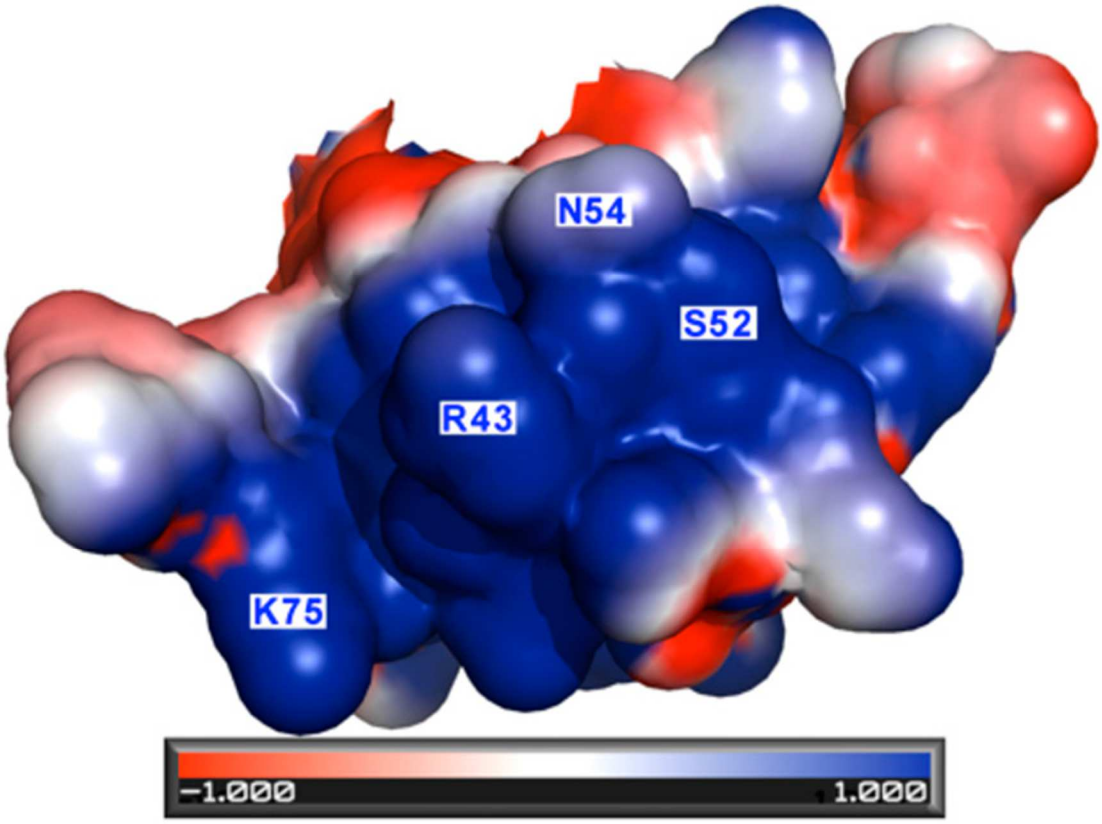


Figure 14. The most basic surface on the GabR winged-helix domain. The electrostatic potential surface of the GabR winged-helix domain is shown. The values −1.000 and 1.000 on the scale bar represent acidic and basic electrostatic potential surface charges, respectively. The orientation corresponds to the wingedhelix domain shown on the bottom in Figure 8. The labeled residues denote amino acids that are conserved between GabR and FadR.

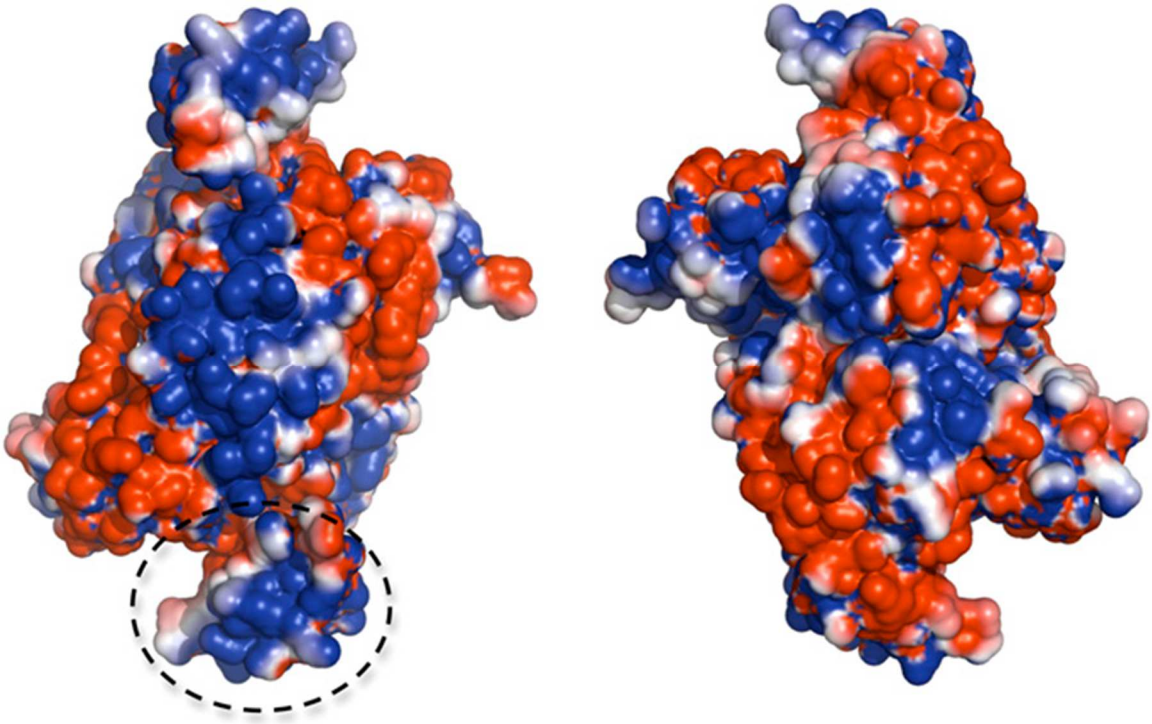


Figure 15. GabR electrostatic surface potential. The electrostatic surface potential representation for the GabR homodimer in two different views. The figure on the Left is in the same orientation as Figure 9. The dotted circle highlights the most basic surface on the winged-helix domain. The figure on the Right is rotated 90° counterclockwise relative to the figure on the Left.

To examine the DNA-binding roles of these conserved residues, we generated a GabR−DNA model by superimposing the structural coordinates of the GabR winged-helix domain onto those of the FadR−DNA complex [Protein Data Bank (PDB) ID code 1H9T [35]]. In the FadR−DNA complex (Figure 16A), two winged-helix domains of a FadR homodimer bind to a 19-bp operator sequence containing two copies of an inverted repeat, TCTGGT, separated only by 3 bp [35]. In contrast, the 47-bp GabR binding site contains two copies of the direct repeat ATACCA separated by 34 bp of DNA. Because we do not know exactly how GabR binds DNA, we only used one copy of the GabR

winged-helix domain for the structural superposition (Figure 16B). The resulting GabR−DNA model (Figure 16B) is identical no matter which FadR monomer we use for the superposition and has several salient features. First, the HTH motif in FadR binds in the major groove of DNA, whereas the wing invades the minor groove (Figure 16A). Residue R35 in helix α2 of the FadR HTH motif makes two H-bonds with the first guanine in the TCTGGT repeat (Figure 16A).

The conserved arginine residue R43 in GabR (Figure 16B)may form similar sequence-specific contacts with one of the guanines in its direct repeat (double-stranded sequence, ATACCA•TATGGT). In the turn of the HTH motif, FadR residue T44 (S52 in GabR) makes H bonds with the phosphate backbone and provides a hydrophobic contact to the central C−G base pair in the 19-mer, whereas residues in the α3 recognition helix, which are typically known to bind the major groove of DNA [30], either contact the central C−G base pair in the 19-mer (T46 in FadR; N54 in GabR) or a phosphate group (T47 in FadR; S55 in GabR) in the major groove. In the FadR wing, residue K67 points between two phosphate groups (Figure 16A). The GabR residue K75 (Figure 16B) could similarly point between two phosphate groups. Because many residues are conserved between FadR and GabR (Figure 17) and because the FadR and GabR DNA repeats have similar sequences, we predict that some of the protein−DNA interactions will be similar between the FadR−DNA and GabR−DNA complexes. However, because the GabR and FadR repeat sequences are dissimilar in type, direction, and spacing, a crystal structure of the GabR−DNA complex is required to eludicate all of the individual pairwise protein−DNA

interactions.

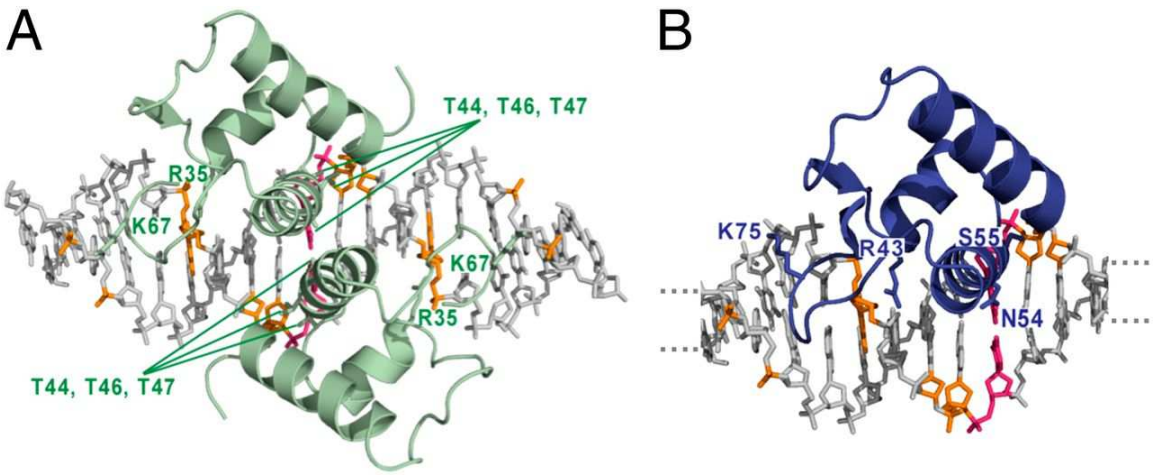
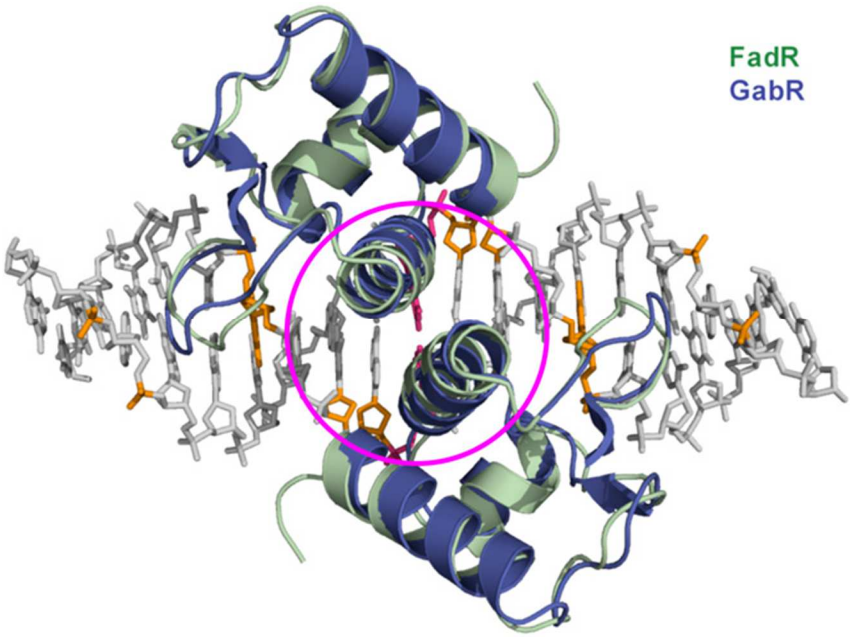


Figure 16. Conserved putative DNA-binding residues in GabR. (*A*) The crystal structure of the FadR−DNA complex is shown (PDB ID code 1H9T) (12). FadR binds DNA as a homodimer. The protein and DNA are shown in cartoon and stick representation, respectively. Side chains of some of the DNA-binding residues are shown as sticks and labeled. The corresponding phosphate groups and nucleotides contacted by the protein side chains are colored in orange. The central base pair in the 19-mer DNA is shown in pink. Only the winged-helix domains of FadR are shown; the acyl-CoA–binding domains are hidden for clarity. (*B*) A model of the GabR winged-helix domain bound to the 19-bp FadR DNA sequence. Again, the central base pair in the 19-mer is colored in pink, although only 12 of the 19 bp are shown for clarity. The phosphate groups and nucleotides putatively contacted by GabR protein side chains are colored in orange. The GabR protein side chains predicted to be involved in DNA binding are shown as sticks and are labeled.

Figure 17. Sequence alignment of the winged-helix domain sequences of *B. subtilis* GabR and *E. coli* FadR. A BLAST alignment of the HTH and wing motifs is shown. The green asterisks highlight DNA-binding residues of FadR that are identical or similar in GabR.

Second, because the two winged-helix domains of FadR interact with each other on DNA (Figure 16A) [35], we examined whether such an interface is possible in GabR. The GabR α3 helix is surface exposed (Figure 8) and has a sequence that is capable of forming an α3−α3 interface similar to that of FadR (Figure 17 and Figure 18). Thus, two copies of the GabR winged-helix domain may be able to interact with each other on DNA under specific circumstances, for instance, when the direct repeats that are 34 bp apart are



brought into close proximity.

Figure 18. Two DNA-binding domains of GabR can interact on DNA through the "α3−α3 dimer interface." A superposition of two molecules of GabR onto the FadR−DNA complex is shown. The α3−α3 interface is indicated by the magenta circle. The acyl-CoA–binding domains of FadR and putative aminotransferase domains of GabR are not shown for clarity. The protein and DNA are shown in cartoon and stick representation, respectively. Side chains of some of the DNAbinding residues are shown as sticks and labeled. The corresponding phosphate groups and nucleotides contacted by the protein side chains are colored in orange. The central DNA base pair is shown in pink. Only the

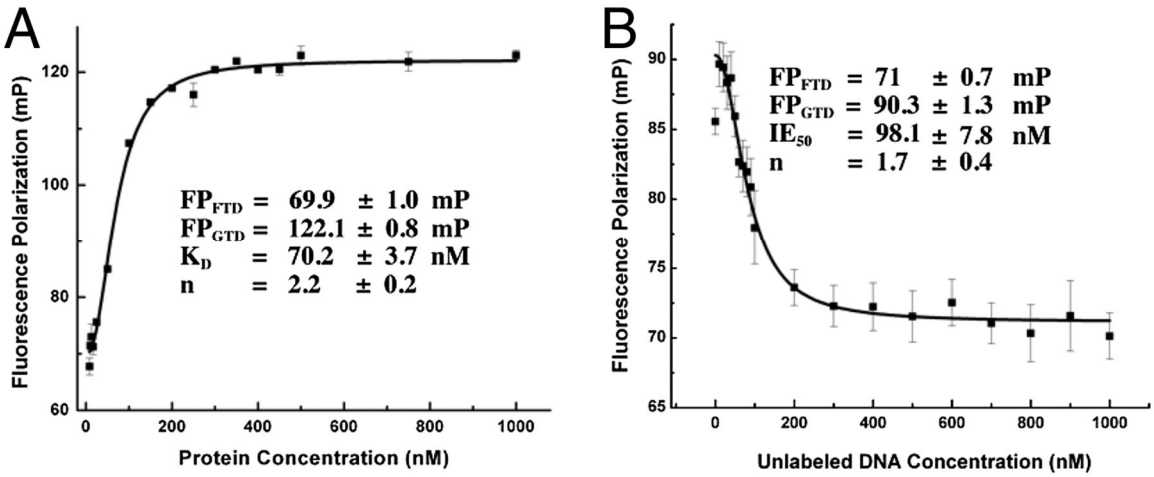
winged-helix domains of FadR are shown. The acyl-CoA–binding domains of FadR are hidden for clarity.

Third, mutating the CC doublets in either of the two GabR direct repeats to GG led to a 10-fold decrease in DNA-binding affinity. Furthermore, mutating both CC doublets to GG completely abolished GabR binding [24]. These observations are consistent with our GabR−DNA model, which suggests that residue R43 of GabR may form sequence-specific contacts with one of the guanines in its direct repeat. As is true for FadR (13) [36], our GabR−DNA model predicts that mutating conserved residues like R43 and K75 may abrogate or interfere with GabR−DNA interactions

GabR Binds DNA with High Affinity

The binding affinity of GabR for a 49-bp DNA containing the 47-bp GabR binding site was determined by fluorescence polarization and competitive binding assays. Previous studies estimated the apparent dissociationExperimental Procedures gabT fragment to be around 1 nM [24]. When a range of GabR protein concentrations were titrated against the 47-bp carboxytetramethylrhodamine (TAMRA)-labeled DNA, we observed a dose-dependent increase in polarization (Figure 19A), indicative of the binding of GabR protein to DNA. The curve was fitted to the modified Hill equation (in Experiments Procedures) and the dissociation constant of the GabR−DNA complex was determined to be 70.2 ± 3.7 nM. A Hill coefficient of 2.2 ± 0.2 was obtained, indicating positive cooperativity in the binding process. To establish specificity of binding and rule out nonspecific effects from the TAMRA label, a competitive binding assay was performed with unlabeled oligonucleotides containing the same sequence as the

TAMRA-labeled DNA (Figure 19B). A preformed mixture of equimolar concentrations of TAMRA-labeled DNA and GabR was titrated with increasing concentrations of unlabeled DNA. The fluorescence polarization decreased with increasing amounts of unlabeled DNA (Figure 19B). The curve was fitted to the Hill equation (Eq. 2 in Experiments Procedures) and the IE50 of unlabeled DNA for GabR was calculated to be 98.1 nM ± 7.8 nM, which is similar to the dissociation constant of the TAMRA-labeled



DNA−GabR complex (Figure 19B). A Hill coefficient of 1.7 ± 0.4 was obtained,

Figure 19. Fluorescence polarization assays to measure DNA binding by GabR. Each data point was measured in triplicate. The average values and errors were plotted and fitted to the Hill equation. The abbreviations used in the figure are as follows: FP_{FTD}, fluorescence polarization signal of free labeled DNA; FP_{GTD}, fluorescence polarization signal of the GabR−DNA complex; K_D, the dissociation constant; n, the Hill coefficient; and IE50, the concentration of unlabeled DNA required to bind 50% of protein. (A) Titration of GabR against 50 nM TAMRA-DNA. The GabR concentrations were varied from 0 to 1,000 nM. Error bars are too small to be displayed if the errors are less than 1/1,000th of the measured value. (B) Competitive binding assay. Titration of unlabeled DNA (0–1,000 nM) against a fixed concentration of preformed GabR-labeled DNA complex (100 nM each of GabR and TAMRA-DNA).

supporting the same type of positive cooperativity as shown in (Figure 19A). Although the precise binding stoichiometry of GabR for DNA is still unknown, the dissociation constants that we determined suggest a high-affinity GabR−DNA complex. The Hill coefficients in both titrations indicate positive cooperativity in the binding of GabR dimers to DNA.

PLP is bound in the C-terminal AT-Fold Domain

Like in other type I aminotransferases [31, 32], the AT-fold domain of GabR has two subdomains, the large subdomain containing the invariant lysine (K312) that forms the Schiff base with PLP and the small subdomain that typically interacts with the α-carboxylate group of the substrate. One molecule of PLP is bound per GabR monomer as an internal aldimine at K312. We next compared the PLP-binding pocket of B. subtilis GabR (Figure 20) with those of E. coli Asp-AT (PDB ID: 2AAT), the best-characterized member of the type I aminotransferase family; and pig liver GABA-AT bound to the substrate analog, γ-ethynyl GABA (GEG) (PDB ID: 1OHY) [32, 37]. The active site residues that facilitate PLP binding are conserved in all three cases. For example, the phosphate group of PLP is situated very close to the N terminus of a nearby helix in all cases and is stabilized by hydrogen bonds from polar residues (T181, T309, and S311 in the case of GabR) to the phosphate oxygens. The salt bridge between an arginine residue (residue R319) and the phosphate group is also conserved in GabR, although the phosphate group has a different orientation compared with those in Asp-AT and GABA-AT. The primary function of PLP is to stabilize the negative charge generated at the Cα atom of the

transition state or external aldimine (Schiff base between an amino acid substrate and PLP) by delocalizing the negative charge through the π-electron system of the cofactor. A conserved aspartate residue [D279 in GabR, D222 in E. coli Asp-AT, D223 in chicken mitochondrial Asp-AT [27], and D298 in GABA-AT; Figs. 5A and 6A] promotes this electron sink nature of PLP by maintaining PLP in the protonated state through interaction with the pyridinium nitrogen.

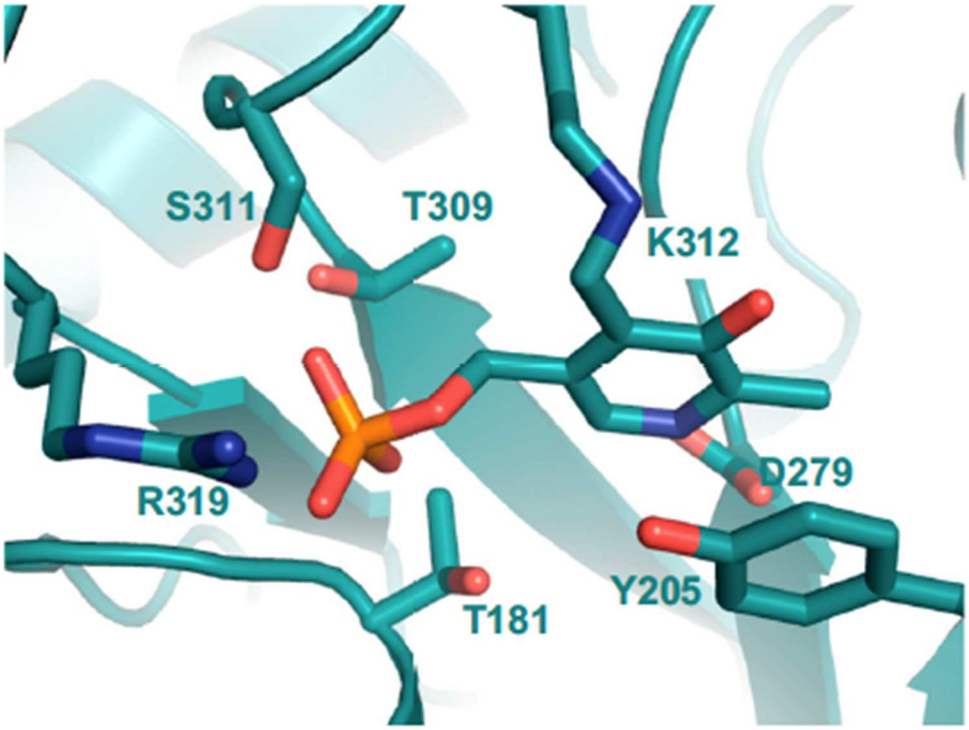


Figure 20. The PLP-binding pocket of B. subtilis GabR. The GabR residues that interact with PLP are labeled.

Is GabR an enzyme?

We were currently unable to obtain crystals of a GabR−PLP−GABA complex, and therefore, we examined the putative GABA-binding site in GabR−PLP to understand whether GabR might catalyze an aminotransferase reaction involving GABA.

Transamination reactions are ping-pong reactions that require two different amino acids to be able to bind in the active site (GABA and glutamate, if GabR were an α-ketoglutarate– dependent GABA aminotransferase). Most aminotransferases use either two conserved arginines or an extended hydrogenbonding network to recognize and bind two amino acids with different shapes and properties [31]. We superimposed the crystal structures of B. subtilis GabR, pig liver GABA-AT−GEG [37], and E. coli Asp-AT [32] to elucidate how different amino acids might bind in the active site of GabR. Even though GabR, GABA-AT, and Asp-AT share the overall type I aminotransferase fold, their polypeptide sequences are highly divergent. Also, the lengths of their secondary structural

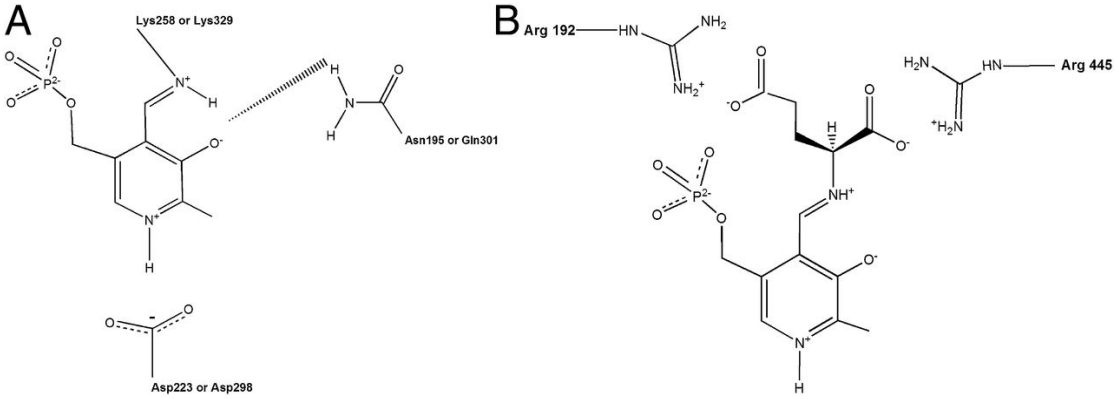


Figure 21. PLP binding in aminotransferases and the arginine switch. (*A*) A schematic of the interactions of the phenolic ring of PLP with surrounding protein side chains. The corresponding residue numbers for Asp-AT (D223, N195, K258) and GABA-AT (D298, Q301, K329) are indicated. The precise locations of the hydrogen atoms have not been observed or measured so far. The H-atom positions indicated in the schematic are therefore only inferred from reaction chemistry. (*B*) A schematic model for the arginine switch in GABA-AT. The model is inferred from the crystal structure of pig liver GABA-AT−GEG complex and adapted. The positions of the two conserved arginine residues in the active site of GABA-AT and their interactions with the α- and γ-carboxylate groups of glutamate are indicated.

elements vary significantly and contain insertions of α-helices and β-strands in different

parts of the structures, making it difficult to obtain secondary structure-based superposition. Therefore, we manually superimposed the pyridine ring of PLP to compare the putative substrate- binding pockets in all cases.

If GabR were a bona fide aminotransferase, then the active site must be able to interact with a keto acid such as α-ketoglutarate. In the active sites of Asp-AT and GABA-AT, two conserved arginines stabilize such a keto acid to subsequently form the related amino acid (Figure 21A)[31]. However, we found two key differences in the "active site" of GabR, compared with these aminotransferases, which may compromise its ability to transaminate GABA. First, the GabR active site may contain only one structurally conserved arginine (R207) that is, however, not conserved by sequence. In Asp-AT and GABA-AT, the equivalent arginine residue (R292* in the other subunit of the dimer in Asp-AT and R192 in GABAAT(Figure 21B), is known to form a bidentate hydrogen bond/ion pair with the γ-carboxylate of a dicarboxylate substrate such as glutamate, whereas the second arginine (R386 in Asp-AT and R445 in GABA-AT; Figure 21B) makes a similar direct contact with the α-carboxylate of the glutamate. Furthermore, GABA-AT employs an arginine switch to distinguish between GABA and glutamate [31, 37]. When GABA binds, R192 stays in place to contact the carboxylate of GABA (Figure 22), which is located in the same position as the γ-carboxylate of glutamate, while R445 moves away to accommodate GABA and engages instead in a salt bridge with a nearby glutamate residue (E270; Figure 22A). Because only the first arginine residue in GabR may be structurally conserved (R207; Figure 22B) while the second arginine is not, this

suggests that GabR may bind GABA but not glutamate or other dicarboxylate substrates. In the absence of conserved arginines, aromatic amino acid aminotransferases and branched-chain aminotransferases use large-scale rearrangements of the active site hydrogen-bond network or hydrophobic cavities lined with polar residues to accommodate two different amino acids in the same site (refs. 9 and 15, and references therein). However, because these features are not preserved in GabR, these alternative substrate-binding mechanisms are unlikely for GabR. The second key difference in the active site of GabR is related to its ability to stabilize the GABA external aldimine. When the amino acid substrate forms an external aldimine, the imine N atom of the amino acid and phenolic O atom of PLP could share an H-atom. Therefore, the phenolic O must be optimally oriented to accept the hydrogen from the imine N during reprotonation and hydrolysis to form PMP and release the deaminated substrate (Figure 21A). In Asp-AT,

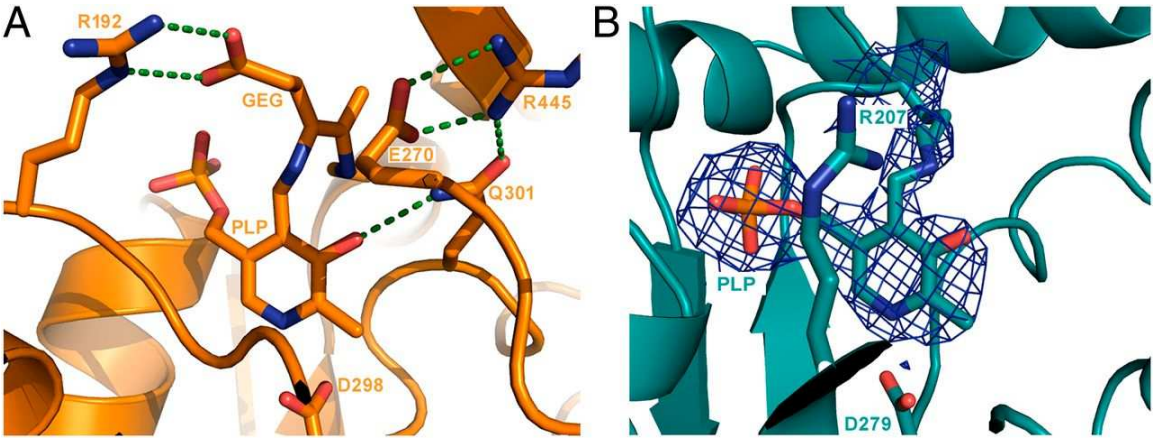


Figure 22. PLP is bound in the AT-fold domain of GabR. (*A*) The substrate-binding pocket in the GABA-AT−GEG structure. The secondary structure elements are shown in cartoon form. Side chains of the key residues that interact with PLP and substrate analog, GEG, are shown in stick representation and labeled. Hydrogen bonds are shown as dotted green lines. (*B*) The putative substrate-binding pocket in the crystal structure of GabR. The figure is drawn as described in *A*. The $2F_o - F_c$ map (blue mesh) around PLP is

shown at 1.2σ level.

an asparagine (N195; Figure 21A) interacts with the phenolic O to hold it in place, and in GABA-AT, a glutamine residue (Q301; Figure 21A) and a water molecule perform the same function. In GabR, there is no such interaction with the phenolic O, implying that the GABA external aldimine might not be optimally oriented for catalysis.

A sequence alignment of 10 MocR/GabR protein sequences from different strains of B. subtilis, E. coli, Pseudomonas aeruginosa, Salmonella enterica, and Streptococcus pneumoniae revealed an overall lack of evolutionary conservation for the arginine switch. Although the residues involved in PLP-binding (equivalent of B. subtilis GabR residues, D279, R319, and K312) are conserved in the examined homologs, we found the residue corresponding to B. subtilis GabR (R207) to be highly variable. In contrast to GabR, all of the PLP-interacting residues and the two substrate-interacting arginines are identical between B. subtilis GabT, a known GABA aminotransferase, and pig liver GABA-AT (Figure 23). Such a contrast indicates the functional divergence of GabR from its evolutionary relatives. Together, these critical differences in the active site architecture of GabR strongly suggest that GabR may bind GABA in a different manner than a true transaminase and may only catalyze the "full" transamination reaction with poor efficiency, if at all. Without the ability to accommodate glutamate effectively, GabR may not be able to complete the second half-reaction and enzymatically regenerate the PLP form of GabR. Consistent with our crystal structure and with previous reports

that GabR does not rescue GabT activity [24], we were unable to measure a GABA-AT activity for GabR in coupled steady-state enzyme assays (described in Experiments Procedures). Also, UV-visible scanning spectrophotometry experiments of GabR in the presence of PLP and GABA did not reveal any of the spectral shifts typically produced by

```
35   KGEGAELYDLDGRRFIDFAGAIGTLNVGHSHPKVVEAVKRQ---AEELIHPGFNVMMYPT  91
     +  G  L D+DG R +D    I ++ +G+SHP +V+ V++    +  +  P   ++
51   ESRGNYLVDVDGNRMLDLYSQISSIPIGYSHPALVKLVQQPQNVSTFINRPALGILPPEN  110

92   YIE-LAEKLCGIAPGSHEKKAIFLNSGAEAVENAVKIARKY--TKRQG-----------  136
     ++E L E L  +AP    +  I +  G+ + ENA K    +  +K +G
111  FVEKLRESLLSVAPKGMSQ-LITMACGSCSNENAFKTIFMWYRSKERGQSAFSKEELETC  169

137  ------------VVSFTRGFHGRTNMTMSMTSKVKPYKFGFGPFAPEVYQAPFPY--YYQ  182
                 ++SF   FHGRT   ++ T     +K     F   +  APFP   Y
170  MINQAPGCPDYSILSFMGAFHGRTMGCLATTHSKAIHKIDIPSFDWPI--APFPRLKYPL  227

183  KPAGMSDESYDDMVIQAFNDFFIA-SVAPETVACVVMEPVQGEGGFIIPSKRFVQHVASF  241
     +     ++  +   ++   D  +      +TVA +++EP+Q EGG    S  F + +
228  EEFVKENQQEEARCLEEVEDLIVKYRKKKKTVAGIIVEPIQSEGGDNHASDDFFRKLRDI  287

242  CKEHGIVFVADEIQTGFARTGTYFAIEHF--DVVPDLITVSKSLAAGLPLSGVIGRAEML  299
      ++HG  F+ DE+QTG   TG ++A EH+  D   D++T SK +     ++G     E
288  SRKHGCAFLVDEVQTGGGSTGKFWAHEHWGLDDPADVMTFSKKM-----MTGGFFHKEEF  342

300  DAAAPGELGGTYAGSPLGCAAALAVLDIIEEEGLNERSEEIGKIIEDKAYEWKQEFP-FI  358
        AP  +  T+ G P        V++II+ E L   +   GK++     + +  +P FI
343  RPNAPYRIFNTWLGDPSKNLLLAEVINIIKREDLLSNAAHAGKVLLTGLLDLQARYPQFI  402

359  GDIRRLGAMAAIEIVKDPDTREPDKTKAAAIAAYANQNGLLLLTAGINGNIIRFLTPLVI  418
       +R  G   + +         PD++    + + A   G++L   G     IRF   LV
403  SRVRGRGTFCSFDT--------PDESIRNKLISIARNKGVML--GGCGDKSIRFRPTLVF  452

419  SDSLLNEGLSIL  430
      D   +  L+I
453  RDHHAHLFLNIF  464
```

Figure 23. BLAST protein sequence alignment. Sequence alignment between B. subtilis GabT GABA aminotransferase (upper sequence; 1–436) and pig liver GABA-AT (lower sequence; 1–472) are shown. The two proteins share 25% sequence identity and 43% similarity between their AT domains. The conserved residues interacting with the substrate carboxylate groups (R192 and R445), the pyridinium N and phenolic O of PLP

(D298 and Q301, respectively), and the invariant Schiff base lysine (K329) are highlighted. The residue numbers in parentheses denote residues from pig liver GABA-AT.

the PLP or PMP intermediates in a PLPdependent transamination reaction. Together, these experiments suggest that GabR may have very poor or no GABA aminotransferase activity under physiological conditions. Although the trifunctional NAD-dependent S. enterica repressor. NadR is an example of a cofactor-dependent protein containing an N-terminal DNA-binding domain and two catalytically active domains [a middle adenylyltransferase domain and a C-terminal nicotinamide ribose kinase [4]], there is also precedence for adaptation of enzymes to regulatory roles in other proteins. For instance, PyrR, a UMP-dependent transcription attenuator for pyrimidine biosynthesis in B. subtilis contains a type I phosphoribosyltransferase (PRT) domain without measurable catalytic activity [38]. The B. subtilis purine repressor PurR has a two-domain architecture like GabR: an N-terminal wingedhelix domain and a C-terminal PRT domain devoid of catalytic activity [39]. It is of course possible that GabR recognizes other substrates or possesses a different PLP-dependent enzymatic activity that may or may not be related to its function as a PLPand GABA-dependent transcriptional activator of gabT. However, so far, no other enzymatic activity for GabR has been reported.

Model for GabR-Mediated Transcriptional Regulation at the gabR and gabT Promoters

Based on the crystal structure of GabR, the sedimentation velocity data and the model for the GabR−DNA complex, we propose three possible modes for GabR binding

and regulation at the gabR and gabT promoters (Figure 24). Because we do not know the precise GabR:DNA binding ratio, we consider different models with one or more GabR dimers bound to DNA. In model 1, two copies of GabR dimers bind to the 47-bp DNA with only one winged-helix domain from each dimer binding to each direct repeat. That the Hill coefficient for the GabR−DNA complex is close to 2 (Figure 19) further suggests that GabR may cooperatively bind DNA as a dimer of dimers in the same way that a dimer of λcII dimers binds its DNA operator [40]. In the λcII−DNA crystal structure, as in model 1, only one monomer from each λcII dimer binds its DNA sequence comprising two direct repeats. Even though the spacings between the direct repeats in the binding sites of λcII and GabR are different (6 and 34 bp, respectively), the cooperative binding of GabR to DNA as a dimer of dimers is still possible. In model 2, only one copy of the GabR dimer binds to the 47-bp DNA and both winged-helix domains of the dimer are involved in DNA binding. In this model, the two winged-helix domains of one GabR dimer may be rearranged and some bending of DNA may occur. The small buried surface area (422 Å2) between the winged-helix domain and AT-fold domains suggests that the position of the wingedhelix domain may be dynamic in

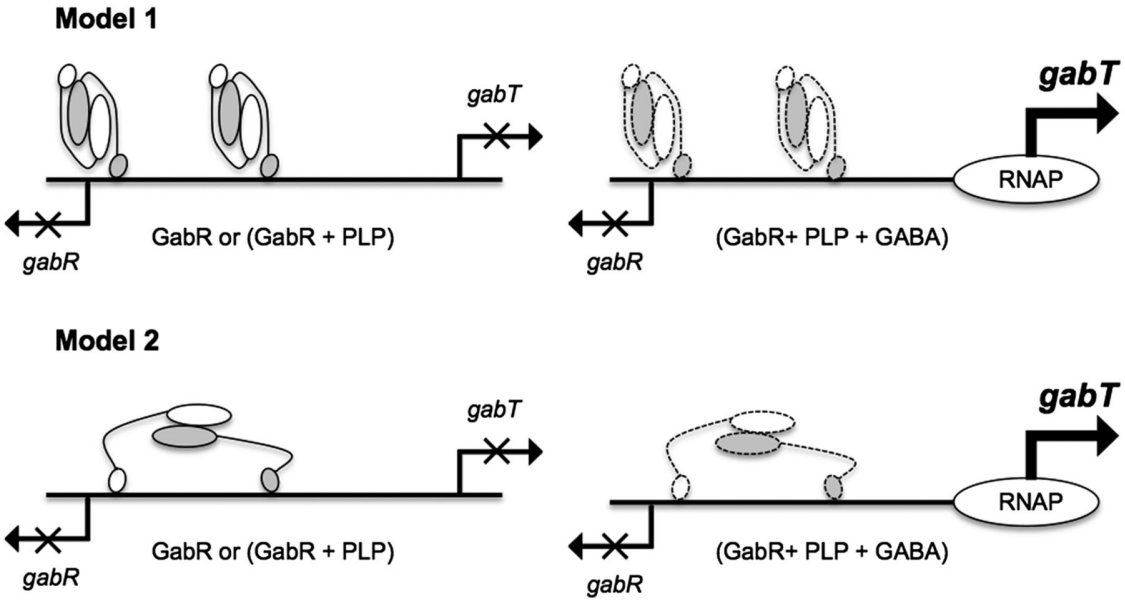


Figure 24. A model for transcriptional regulation at the *gabR* and *gabT* promoters. Different models for GabR binding in the *gabR–gabT* regulatory region are depicted. The GabR homodimer is represented as a cartoon with one monomer shown in white and the other colored gray. The larger oval represents the AT-fold domain and the smaller oval represents the winged-helix domain. Putative conformational changes in the presence of GABA are indicated by symbols with dashed lines.

solution and that the proposed domain movements are possible using large-scale movements facilitated by the long linker. We recognize that other viable models still exist. For example, the second winged-helix domain of a GabR dimer may bind to unidentified cis binding sites located at larger distances from the gabR−gabT locus to form large DNA loops. Intriguingly, a large basic surface at the AT−AT interface begs the question whether this surface is somehow also involved in DNA binding (Fig. S8). These models suggest that GABA may bring about conformational changes in GabR or the GabR−DNA complex, and that these changes may be necessary to activate gabT transcription, for instance, by recruiting RNA polymerase to the gabT promoter (Figure

24). Crystal structures of GabR−DNA complexes with and without GABA may help unravel some of these unknowns.

Biological Implicaions

Our studies suggest that GabR may not possess "full" GABA aminotranferase activity, although it requires PLP and GABA to activate transcription at the gabT promoter. It is likely that the primary function of GabR in B. subtilis is to regulate transcription at the gabR and gabT promoters and that the AT-fold domain of GabR has adapted to primarily serve as an effector binding domain. How the AT-fold domain of GabR assists the wingedhelix domain in transcriptional regulation and how GABA mechanistically affects gabT activation by GabR still remain to be determined.

In summary, these studies were published in the in the Proceedings of the National Academy of Sciences of the United States of America (PNAS): Raji Edayathumangalam, Rui Wu, Yuguang Wang, Roman Garcia, Wei Wang, Cheryl Kreinbring, Jingling Liao, Todd Stone, Quyen Q. Hoang, Boris Belitsky, Gregory A. Petsko, Dagmar Ringe, and Dali Liu. (2013) 110 (44), 17820-17825.(Equal contribution first author)

Materials and Methods

Protein Expression and Purification. The full-length coding region of the Bacillus subtilis gabR gene (residues 1–479) was subcloned from a pBAD construct into pETite C-His vector (Lucigen), which encodes an inducible T7 expression system. The C-terminally His6-tagged protein was expressed in Escherichia coli strain BL21 (DE3).

Transformed cells were grown in LB media containing kanamycin at 37 °C to OD600 of □0.6 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37 °C for 4 h. Cells were harvested, resuspended in buffer containing twofold PBS buffer and 10 mM imidazole (pH 7.5), and sonicated. Insoluble material was removed by centrifugation. The GabR-His6 protein (molecular weight, 56 kDa) was purified to homogeneity by affinity chromatography Ni-NTA (GE Healthcare) and gel filtration (HiLoad 16/60 Superdex 200 pg; GE Healthcare) in buffer containing twofold PBS (274 mM NaCl, 5.4 mM KCl, 16 mM Na2HPO4, 2.92 mMKH2PO4, pH 7.4), 475 mM imidazole (pH 7.5), 1 mM pyridoxal 5′-phosphate (PLP), and 5% (vol/vol) glycerol. The protein requires high salt and imidazole concentrations to stay soluble without precipitation. Alternatively, the presence of 5% (vol/vol) glycerol and 0.5% Nonidet P-40 in the buffer helps to keep GabR soluble without the need for NaCl and imidazole. Purified GabR-His6 protein was concentrated (Centricon Plus centrifugal filter units; Millipore) and stored at 4 °C for short term (□4 wk) or flash frozen in □50-μL aliquots in liquid nitrogen after the addition of 15% (vol/vol) glycerol and stored at −80 °C for long-term use. The concentrated sample of purified GabR was pale yellow in color, indicating that PLP was bound to the protein.

Analytical Ultracentrifugation Analysis. The oligomeric state of GabR−PLP in solution was analyzed by sedimentation velocity using a Beckman-Coulter Optima XL-I ultracentrifuge. The experiments were conducted at low protein concentrations (30, 50, and 80 μg/mL) in the presence of 0.5 mM PLP and in the presence and absence of 20 mM

GABA. In each experiment, □450 μL of the protein solution was loaded into a sector-shaped centrifugation cell at concentrations corresponding to OD230 values of 0.3, 0.5, and 0.8. The samples were spun at 48,000 rpm (Beckman AN60Ti Rotor) for 5 h with radial scans capturing the boundary distribution every 2 min until the sample had almost completely pelleted. Boundary profiles from sedimentation velocity scans were collected and analyzed using the UltraScan II software suite. Initial model-independent van Holde–Weischet analysis of the data clearly showed no evidence of mass action over the concentration range investigated, thereby confirming an absence of dynamic equilibrium between species. Therefore, all data were fit to a 50-iteration global genetic algorithm–Monte Carlo (GA-MC50) model, which showed a narrow and randomly distributed residual, indicative of good data quality and modelfitting.

Fluorescence Polarization DNA-Binding Assay. Fluorescence polarization data for the GabR–DNA complex were obtained using an EnVison 2102 Multilabel Plate Reader (Perkin-Elmer). Two 49-bp complementary single-stranded DNA oligonucleotides containing the 47-bp GabR binding region (5′ TCTGATACCATCAAAAAGTTATAATTGGTACTTTTCATCATACCAAAGA-3′) were purchased from Integrated DNA Technologies. The oligonucleotides were either unlabeled or labeled at the 3′ end with the fluorophore, TAMRA. The double-stranded oligonucleotides were prepared by annealing a mixture containing equimolar concentrations of the two single-stranded oligonucleotides. For the fluorescence polarization assay, varying concentrations of the GabR protein was titrated against a

fixed concentration of TAMRA-labeled DNA (50 nM). GabR protein was serially diluted in the DNA binding buffer [20 mM Tris·HCl, pH 8.0, 50 mM KCl, 2 mM MgCl2, 5% (vol/vol) glycerol, 1 mM EDTA, 1 mM DTT, and 0.05% Nonidet P-40]. The TAMRA-DNA (50 nM) was mixed with the serially diluted GabR protein (final concentration, 0–1,000 nM) and 100 μL of each sample mixture was loaded into the individual wells of a 384-well microplate. In the competitive binding assay, varying concentrations of unlabeled DNA were titrated against a fixed concentration of preformed GabR-labeled DNA complex (100 nM GabR and 100 nM TAMRA-DNA). The fluorescence polarization (FP) signals of the GabR−DNA mixtures were measured with fixed excitation (531 nm) and emission (595 nm) wavelength filters at room temperature. Fluorescence intensity remained constant throughout the polarization measurements. The fluorescence polarization signal for each sample was calculated and plotted against the GabR protein concentrations in the fluorescence polarization assay and against the concentrations of unlabeled DNA in the competitive binding assay. Each data point is the average of three independent measurements. The titration plot for GabR versus TAMRA-DNA was fitted to a modified Hill equation (Eq. S1) in ORIGIN software to calculate the dissociation constants and Hill coefficient. For the competitive binding assay, the IE50 value and Hill coefficient were obtained by fitting the data points to the Hill equation. (Eq. S2).

$$(FPGTD - FPFTD) = [TAMRA\text{-}DNA]^n / K^n_D + [TAMRA\text{-}DNA]^n; \quad [1]$$

where FPGTD represents FP signal when TAMRA-DNA is saturated with GabR; FPFTD

represents FP when all TAMRADNA is free in solution without any GabR bound; KD is the dissociation constant for the GabR–TAMRA-DNA complex; and n is the Hill coefficient.

$$(FPGTD - FPFTD) = [\text{Unlabeled DNA}]^n / (IE^n_{50} + [\text{Unlabeled DNA}]^n), \quad [2]$$

where FPGTD, FPFTD, and n are the same as in Eq. S1; IE50 represents the concentration of unlabeled DNA needed to bind 50% of protein.

Crystallization. Crystals of GabR were obtained by the hangingdrop method using protein at 5 or 10 mg/mL The initial sparse matrix screen was set up with the Gryphon crystallization robot (Art Robbins Instruments) to obtain initial crystallization hits, which were further optimized manually to obtain crystals with the best size and morphology. The optimized crystallization solution contained 33% (vol/vol) PEG400, 200 mM calcium acetate, and 200 mM imidazole, pH 7.5. The hanging drops were incubated at 20 °C or room temperature. Crystals appeared in 5 d and grew to maximum size in 1–2 wk. Crystals with good morphology and size were picked directly from the hanging drops and flash-cooled in liquid nitrogen. The heavy metal derivatives were prepared by soaking crystals in 10 mM ethylmercury phosphate for 1 h before flash- cooling in liquid nitrogen. The heavy metal soaking condition also included 1 mM PLP.

Data Collection and Processing. Single-wavelength anomalous dispersion (SAD) data were collected at beamlines 19-ID and 19-BM at Advanced Photon Source, Argonne National Laboratory. X-ray diffraction data for the mercury derivative of holo-GabR (Table 1, Dataset 1) and for holo- and apo-GabR native crystals (Table 1, Datasets

2 and 3) were collected at the wavelengths of 1.0039 and 0.9792 Å, respectively. The datasets were indexed, integrated, and scaled with either the HKL package [41] or iMosflm and Scala programs in the CCP4 program suite [42]. The datasets were all scaled in the space group P212121 to resolutions of 3.1, 2.7, and 2.55 Å, respectively. The data statistics are reported in Table 1.

Table 1. Data collection and refinement statistics for full length GabR

Datasets	(1) GabR-PLP, mercury derivative	(2) GabR-PLP, native	(3) apo-GabR, native
Wavelength (Å)	1.0039	0.9792	0.9792
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution range (Å)	30-3.1 (3.15-3.1)	50-2.71 (2.76-2.71)	49.3-2.55
Unit cell (a b c) (Å)	98.1 101.1 211.2	97.2 101.3 211.2	99.9 101.4 213.2
Total unique reflections	39106	57413	68270
Multiplicity	6.7 (6.8)	4.8 (3.8)	5.8 (5.6)
Completeness (%)	100 (100)	99 (97.9)	95.8 (90.7)
Linear R_{merge}[a] (%)	10.7 (71.6)	0.102 (ND)	11.7 (68.7)
I/σ	21.9 (2.9)	18.7 (1.0)	9.4 (2.2)
Average Wilson B-factor	67.7	59.9	38.7
R_{cryst}[b]	0.2124		0.1947
R_{free}[c]	0.2556		0.2461
Number of atoms	15247		15514
Macromolecules	15108		15139
Ligands	99		73
Water	40		300
RMSD[d] bonds (Å)	0.005		0.010
RMSD[d] angles (°)	1.02		1.405
Ramachandran favored (%)	94.0		89.8
Ramachandran outliers (%)	1.5		3.5
Average B-factor (Å^2)	82.7		47.4
Macromolecules	82.6		63.0
Solvent	55.2		50.5

The values for the highest resolution bin are in parentheses.

[a]Linear $R_{merge} = \Sigma|I_{obs}-I_{avg}|/ \Sigma I_{av}$

[b]$R_{cryst} = \Sigma|F_{obs}-F_{calc}|/ \Sigma F_{obs}$

[c]Five percent of the reflection data were selected at random as a test set and only these data were used to calculate R_{free}.

[d]RMSD, root mean square deviation.

ND, Not Determined/Calculated by the program

Sequence and Structure Analyses. GabR homologs were identified in the sequence database using BLAST ; multiple sequence alignment using BLAST; conserved domain searches using CDD (4) [34]; buried surface area from PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver); and structural similarity using DALI [43]. Secondary structure superimpositions were done in Coot. Electrostatic surface potentials were calculated with PyMOL (The PyMOL Molecular Graphics System, version 1.5.0.4; Schrödinger). Figs. 3, 4, 6, and 9, and Figs. S2, S4, S5, S7, and S8 were drawn with PyMOL.

Structure Solution, Model Building, and Refinement. Automated substructure solution by AutoSol in PHENIX [44] found 15 heavy atom peaks in the mercury dataset (Table 1, Dataset 1) and identified fourfold noncrystallographic symmetry (NCS) using superposition of the mercury sites. AutoBuild in PHENIX automatically built □80% of the residues. The initial phases were extended and refined to 2.7 Å using the native holo-GabR dataset (Table 1, Dataset 2) to yield improved maps and further improved by

fourfold NCS averaging and density modification. The remainder of the model was built manually in Coot [45] guided by the structure of Thermococcus profundus multiple substrate aminotransferase [Protein Data Bank (PDB) ID code 1WST], a homolog obtained from the FFAS03 server (http://ffas.burnham.org/ffas-cgi/cgi/ffas.pl). Several loops accounting for gaps in the model were placed and refined by a combination of manual building and the fit_loops routine in PHENIX. Next, MR-SAD using the model and the mercury dataset identified five more mercury sites and led to even better maps. Several iterative rounds of manual model building and refinement using the maximum-likelihood amplitude targets in PHENIX and Rosetta using simulated annealing (in the initial stages) and automatic NCS restraints were required to obtain the final model. The initial model (□80% complete) had Rcryst and Rfree values of 0.393 and 0.484, respectively, for all data between 45.0 and 2.7 Å. The Rcryst and Rfree values of the final model (98% complete) are 0.229 and 0.265, respectively. The final model of the GabR−PLP complex consists of two homodimers of GabR in the asymmetric unit with □470 residues out of 479 accounted for in each monomer. The final model also contains 1 molecule of PLP per monomer, 16 molecules of ethylmercury, and 52 water molecules. The quality of the final model is summarized in Table S1. This model has been deposited in the PDB and is available with ID code 4N0B. The native dataset of apo-GabR (Table 1, Dataset 3), which was processed to a resolution of 2.55 Å, was phased by molecular replacement using the final model of GabR as the input model.

The Rfactor and Rfree values of the final model from the native dataset are 0.195

and 0.246, respectively. Other model statistics are summarized in Table S1. This model has been deposited in the PDB with ID code 4MGR. The density for PLP is lacking in this structure, as expected. We interpreted a difference density peak in the PLP-binding pocket as imidazole (Figure 10) because imidazole is a major component in the protein purification and crystallization buffer.

GabR Enzyme Assay. Activity was measured using varying concentrations of GabR protein in an assay buffer containing 100mM Tris·HCl, 5 mM DTT, 100 mM KCl, 100 μM PLP, 20 mM GABA, 20 mM α-ketoglutarate, 500 μM NADP+, and 3 units of aldehyde dehydrogenase (ADH). In the coupled assay, ADH consumes the product of the first half-reaction (succinic semialdehyde) to produce succinate. The coupled reaction is measured spectrophotometrically (absorbance change at 340 nm) from the concomitant reduction of the cofactor, NADP+ to NADPH.

CHAPTER THREE

PLP AND GABA TRIGGER GABR-MEDIATED TRANSCRIPTION REGULATION IN BACILLUS SUBSILIS VIA EXTERNAL ALDIMINE FORMATION

Research Significance

The *Bacillus subtilis* protein GabR activates the biosynthesis of glutamate from γ-aminobutyric acid (GABA) at the transcription level upon binding to pyridoxal-5'-phosphate (PLP) and GABA. We herein report that the external aldimine formed between PLP and GABA is apparently responsible for triggering the GabR-mediated transcription activation. A crystal structure of the GabR effector-binding/oligomerization (Eb/O) domain (residues 88-479) in complex with both PLP and GABA revealed the formation of a PLP-GABA Schiff base. Details of the GABA binding site in the Eb/O domain structure suggest that binding monocarboxylic γ-amino acid such as GABA should be preferred over dicarboxylic acid ligands. A reactive GABA analog, *(S)*-4-amino-5-fluoropentanoic acid (AFPA), was used as a molecular probe to examine the reactivity of PLP in both GabR and a homologous aspartate aminotransferase (Asp-AT) from *E. coli* as control. The comparison between the complex structures of Eb/O-PLP-AFPA and Asp-AT-PLP-AFPA revealed that GabR is incapable of facilitating further steps of the transamination reaction after the formation of the external aldimine. Results of *in vitro* and *in vivo* assays using full-length GabR support the conclusion that AFPA is a

agonistic ligand capable of triggering GabR-mediated transcription activation via formation of an external aldimine with PLP

Introduction

Bacterial transcription regulators are often directly responsive to the changes in metabolite concentration [15, 16]. GabR responds to the increase of GABA concentration in bacterial cells [25] and upregulates glutamate regeneration from GABA. Belonging to the MocR/GabR subfamily in the GntR family of bacterial transcription regulators [16, 25]. GabR-type regulators characteristically have a large C-terminal Eb/O domain, which is homologous to the type I aminotransferases [31], and a small N-terminal "winged-helix-turn-helix" (wHTH) DNA-binding domain [26]. In *B. subtilis*, GabR represses its own transcription [24, 25]. With both PLP and GABA bound, GabR activates the transcription of two genes, gabT and gabD, which encode the enzymes GABA aminotransferase (GABA-AT) and succinic semialdehyde dehydrogenase (SSA-DH) respectively, forming a pathway for glutamate regeneration from GABA [24, 25].

Previously, we have reported the crystal structures of full-length GabR as a "head-to-tail" dimer [46], in which electron density supported the existence of PLP bound to K312 of the Eb/O domain in the form of an internal aldimine. Although the PLP-binding site of GabR is largely conserved compared to that of type I aminotransferases, no aminotransferase activity has been detected for GabR [46]. To our knowledge, the existence of PLP in MocR/GabR- type transcription regulators is the only case known in nature where PLP functions as an effector in transcription regulation, including the other reported case of a MocR regulator PdxR that controls the biosynthesis of PLP in bacteria

[47-49]. A previous report on GabR has shown that a "second metabolite pair" of pyridoxamine-5'-phosphate (PMP) and succinic semialdehyde (SSA) could also cause transcription activation upon binding to GabR [24]. The first half of the "Ping-Pong" transamination catalyzed by GABA-AT (Schemes 1 and 2) in fact converts PLP and GABA to PMP and SSA, reversibly [50]. It was hypothesized that certain step(s) of the aminotransferase reaction might be required to trigger GabR-mediated transcription activation [24]. Such hypothesis has not been proven, and the chemical identity of the GabR-PLP-GABA complex that results in transcription activation remained unknown. Using X-ray crystallography, NMR spectroscopy and biological assays, an external aldimine species formed between PLP and GABA when both bound to the effector binding site of GabR has been identified and is apparently responsible for triggering GabR-mediated transcription activation.

Crystallography of GabR Eb/O Domain and Asp-AT Complexed with Ligands

Three complex structures with PLP-ligand adducts bound were obtained through co-crystllization. X-ray crystal structures were determined and refinment of all three complex structures converged with good statistics (Table 2). The Schiff base adducts were built into the electron density maps (*fo-fc*) in the final rounds of refinement. The formed adduct was confirmed by the simulated annealing omit difference maps (*fo-fc*) generated in Phenix [51]. The final models for EB/O-PLP-GABA, EB/O-PLP-AFPA and Asp-AT-PLP-AFPA have been deposited in the Protein Data Bank (PDB) with accession codes XXXX, YYYY and ZZZZ, respectively. The full length GabR was truncated to its

Eb/O domain (residues 88-479), expressed, purified and co-crystallized with PLP and GABA. The complex structures Eb/O-PLP-GABA and Eb/O-PLP-AFPA were solved with molecular replacement and refined at 2.23 Å resolution and 2.25 Å resolution, respectively. The complex structure of Asp-AT-PLP-AFPA was solved and refined at 1.53 Å resolution. The overall structure of the Eb/O domain retains the same "head-to-tail" homodimer architecture as those in the full-length GabR and type I aminotransferases (Figure 25).

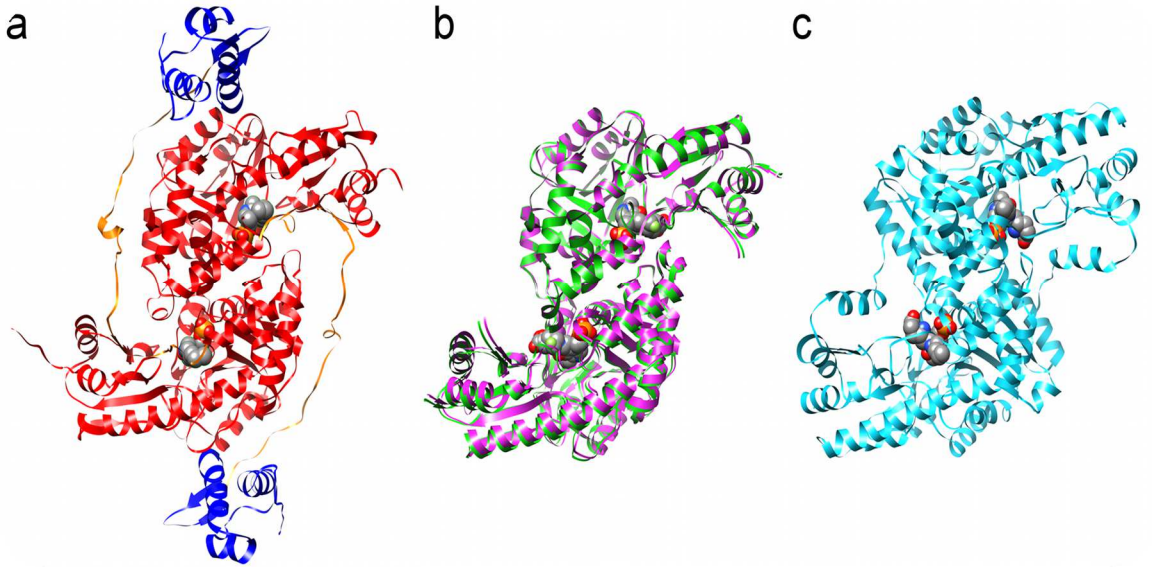


Figure 25. Head-to-tail-dimers of the full-length GabR (PDB code: 4N0B), Eb/O domain of GabR and Asp-AT. Bound PLP's are shown in spheres. a) In the full length GabR, the Eb/O domains are in red, the flexible linker regions are in orange, and the DNA-binding domains are in blue. b) An overlay of two dimeric structures of the GabR Eb/O domain in complexes with GABA (magenta) and AFPA (green). c) The dimeric structure of E. coli Asp-AT (PDB code 1SFF, cyan).

Table 2. Data collection and refinement statistics for GabR Eb/O domain

Protein	**GabR-EBO/GABA**	**GabR-EBO/AFPA**	**Asp-AT/AFPA**
Data processing			
Space group	$C222_1$	$C222_1$	$C222_1$
Cell dimension			
α, β, γ (°)	90.0; 90.0; 90.0	90.0; 90.0; 90.0	90.0; 90.0; 90.0
a, b, c (Å)	92.4; 128.9; 65.1	96.6; 127.5; 67.6	84.9; 154.9; 79.3
Resolution (Å)	49.2 - 2.23	50.8 - 2.25	29.0-1.53
[a]R_{merge} (%)	7.5 (46.4)[b]	8.3 (64.4)[b]	6.8 (68.7)[b]
I/σ (I)	29.3(4.7)[b]	12.6(1.2)[b]	34.0(1.7)[b]
Completeness (%)	99.8(97.9)[b]	96.2(78.4)[b]	98.2(83.4)[b]
Multiplicity	5.5 (4.9) [b]	3.1 (2.5) [b]	12.2 (5.6) [b]
No. Reflections	248492	132836	1914313
No. Unique Reflections	19166	19311	147549
Refinement			
[c]R_{work}/[d]R_{free} (%)	17.44/21.07	18.82/23.70	14.88/16.55
No. of Atoms			
Protein	2921	2292	3465
Ligand	22	24	32
Water	136	182	510
B-factors(Å^2)			
Protein	40.41	42.24	23.72
Ligand (Lowest/Highest)	28.60/41.18	24.80/43.97	13.05/27.80
[e]RMSD (from ideal values)			
Bond length (Å)	0.005	0.003	0.007
Bond angle (°)	0.862	0.604	1.128
Ramachandran plot			
Most Favored (%)	96.69	96.40	97.17
Allowed (%)	3.04	3.32	2.61
Outliers (%)	0.28	0.28	0.22

[a] $R_{merge} = \Sigma|I_{obs}-I_{avg}|/ \Sigma I_{avg}$
[b] The values for the highest resolution bins are in parentheses.
[c] $R_{work} = \Sigma|F_{obs}-F_{calc}|/ \Sigma F_{obs}$.

[d] Five percent of the reflection data was selected at random as a test set, and only these data were used to calculate R_{free}.
[e] RMSD, root mean square deviation

Details of the Effector Binding Site in the GabR Eb/O Domain and Schiff Base Formation between PLP and GABA

In the effector-binding pocket of the Eb/O domain, the omit map difference electron density (*fo-fc*) revealed the existence of a Schiff base formed between PLP and GABA Figure 26a). Three basic residues, R207, R430 and H114, from the same subunit interact with the carboxylate of the GABA moiety. In a complex structure of GABA-AT-PLP-aminooxyacetate (AOA) (Figure 26a), the PLP-AOA adduct mimicks the PLP-GABA external aldimine, whose carboxylate interacts with one of the conserved arginine

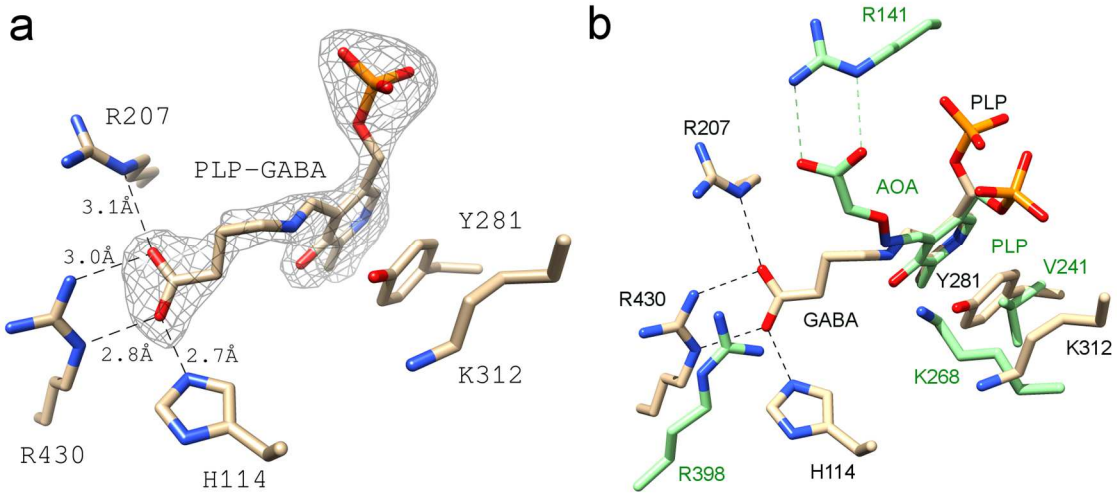
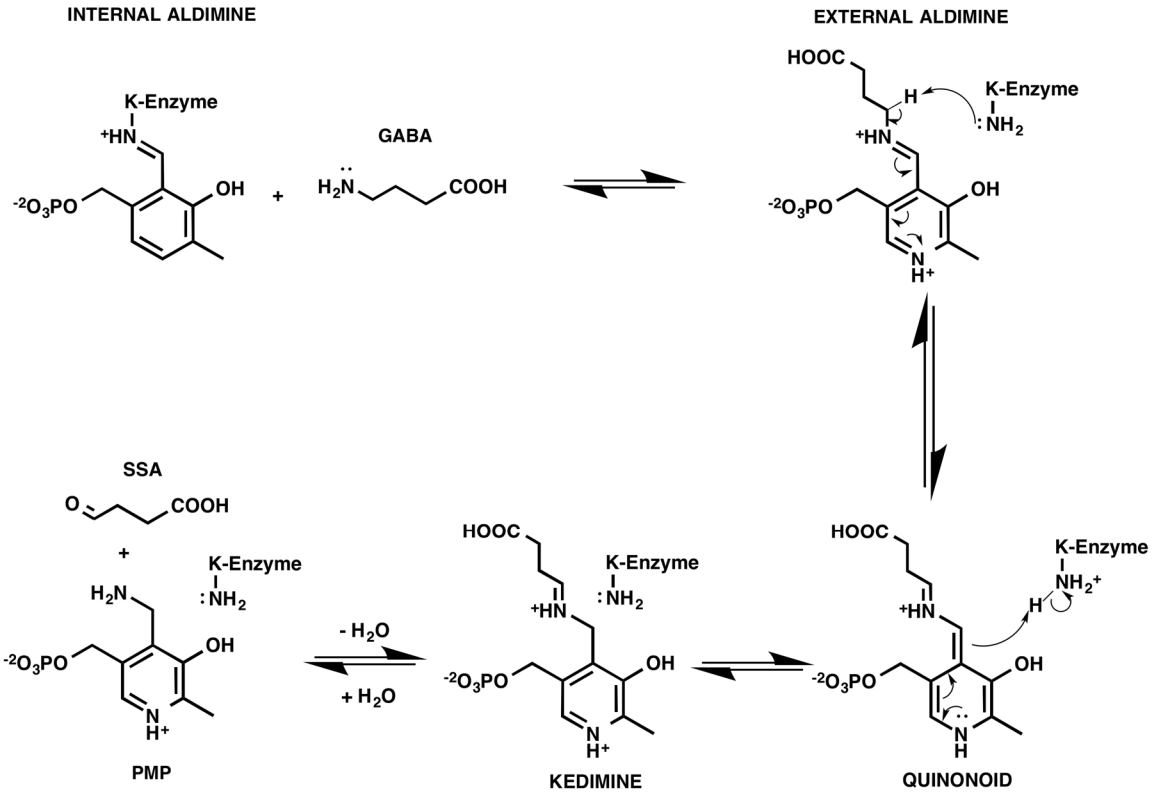


Figure 26. PLP and GABA bound to the truncated Eb/O domain. a) The simulated annealing omit map (fo-fc) is shown around the PLP-GABA Schiff base at 3σ. b) Eb/O-PLP-GABA in comparison to GABA-AT with aminooxyacetate (AOA) bound to PLP mimicking the PLP-GABA external aldimine (green, PDB code: 1SFF). transamination reaction (Scheme 1).

residues (R141) [52]. Compared with the homologous type I aminotransferases such as GABA-AT (Figure 26B), GabR is missing one (R141 in GABA-AT) of the two conserved arginines, which are involved in binding to dicarboxylic acid substrates [37, 52], such as *L*-glutamate or 2-ketoglutarate, in the second-half of a typical

This observed structural difference suggests that GabR has evolved to preferentially recognize monocarboxylic acids, such as GABA, and orient the ligand in favor of forming a Schiff base with PLP.



Scheme 1. The Mechanism of First Half of the GABA-AT Reaction.

Interestingly, the preserved basic residues in GabR are not aligned with GABA-AT residue R141, which is proposed to interact with the GABA carboxylate [33, 52]. While

the aminooxyacetate (AOA)-PLP adduct, which mimics the GABA-PLP in the *E. coli* GABA-AT structure, binds to the active site in an anti conformation, the PLP-GABA Schiff base in GabR is in a gauche conformation (Figure 3b). The dihedral torsional angle of C2-C1-O-N in the PLP-AOA "structural mimic" is measured to be 156.7° (13.3° off from an ideal anti-conformation) in UCSF Chimera. In the complex structure of Eb/O-PLP-GABA, the GABA moiety displays a gauche conformation (Figure 27).

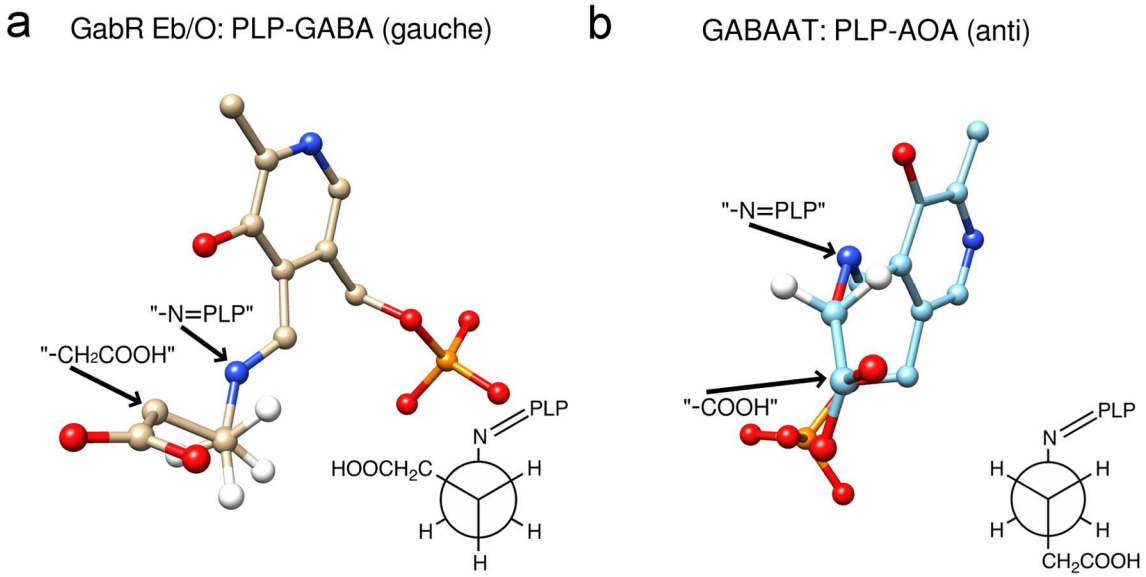


Figure 27. GABA Moiety Displays a Gauche Conformation. (a) The GABA bound in the Eb/O domain is in a gauche conformation. The PLP-GABA external aldimine is shown in stick and ball form. The Cγ atom is placed (hidden) behind the Cβ atom, in the same orientation as the Newman projection for the dihedral torsional angle Cα-Cβ-Cγ-N in the PLP-GABA. A Newman projection of a PLP-GABA Schiff base is shown in a gauche conformation. (b) The aminooxyacetate (AOA)-PLP imine bound in *E. coli* GABA-AT (PDB code: 1SFF) mimicking the PLP-GABA Schiff base was shown in ball and stick form. The O atom is placed (hidden) behind the C1 atom in the same orientation as the Newman projection for the dihedral torsional angle C2-C1-O-N in the PLP-AOA. A Newman projection of a PLP-GABA Schiff base is shown in an anti-conformation.

The dihedral torsional angle of Cα-Cβ-Cγ-N in PLP-GABA Schiff base was measured to be 77.3° (17.3° off from an ideal gauche conformation) in UCSF Chimera. Compared to the anti conformation, the gauche conformation is of higher energy and a relatively minor conformation for GABA molecule dissovled in an aqueous environment in the cell. Herein, we speculate that GabR's preference for a gauche conformation could be an evolutionary outcome to sense the GABA's concentration buildup instead of detecting the basal GABA level inside the cell. Alternatively, we speculate that the chemical strain could be released and provide extra energy for GabR's action after or during transcription activation.

In all PLP-dependent enzymes, a conserved lysine forms an internal aldimine with PLP in a resting form and acts as a key catalytic residue during catalytic reactions, such as in the first half of the "Ping-Pong" transamination [31] (Scheme 1). The conserved K312 [46](Figure 28) was previously proven to be essential for GabR's function, although the catalytic capacity to facilitate transamination was never detected for GabR [24]. In the Eb/O-PLP-GABA complex structure (Figure 2a), Y281 is observed to block K312 from accessing the PLP-GABA Schiff base, preventing the reformation of the internal aldimine as well as further transamination-like catalysis, in which the freed catalytic lysine would deprotonate the carbon adjacent to the nitrogen atom of GABA (Scheme 1) [33, 37]. As a result, GabR is unlikely to catalyze the first half of the transamination reaction shown in Scheme 1. As we previously reported [46], in the absence of GABA, Y281 is found in a different orientation, allowing internal aldimine formation (Figure 28). The observed

obstruction by Y281 supersedes the previous speculation that GabR's inability to stabilize the external aldimine might be the reason for no observed transaminase activity[46].

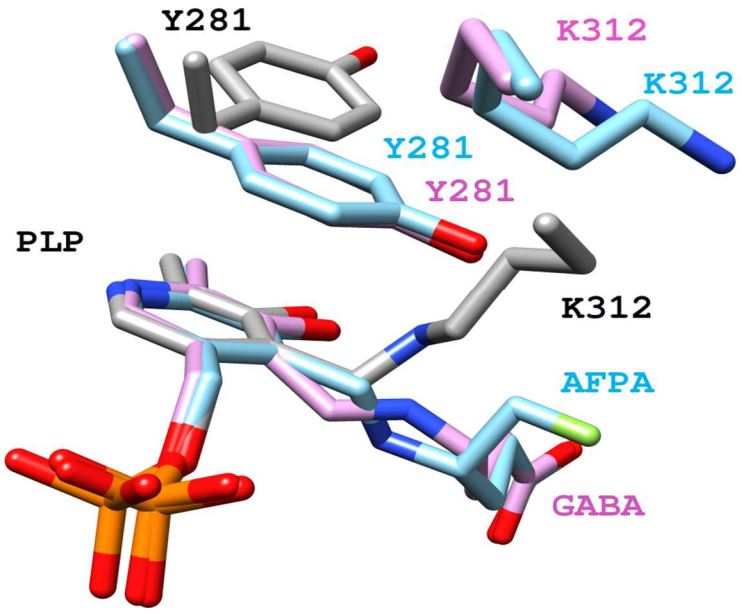
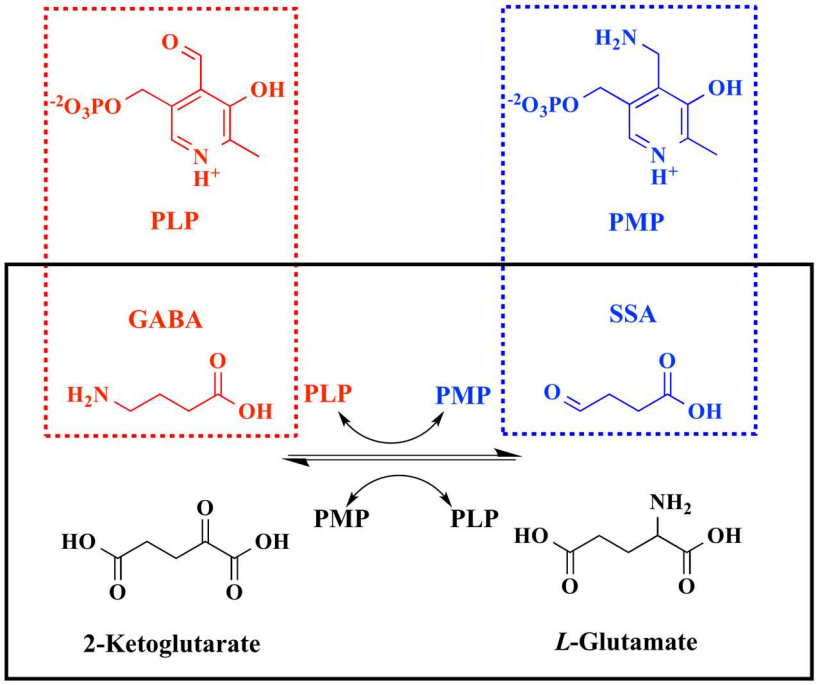


Figure 28. Comparison of the external aldimine vs. internal aldimine in GabR. Schiff bases, K312, and Y281 are shown in stick form. Labels and carbon atoms for internal aldimine (PDB code: 4N0B) are in black and grey, respectively. Labels and carbon atoms for Eb/O-PLP-GABA are in magenta. Labels and carbon atoms for Eb/O-PLP-AFPA are in cyan.

External Aldimine as the Final Adduct of GabR-Mediated Reaction

The Eb/O-PLP-GABA complex structure at 2.23 Å resolution could not allow us to distinguish between external aldimine and ketimine (Scheme 1) for the chemical identity of the PLP-GABA Schiff base observed. Therefore, we used a mechanism-based inactivator, (*S*)-4-amino-5-fluoropentanoic acid (AFPA), which was designed to

irreversibly inactivate PLP-dependent aminotransferases [53], as a molecular probe. A previously proposed inactivation mechanism of an aminotransferase, mammalian GABA-AT, by AFPA is shown in Scheme 2. Once a typical aminotransferase reaction progresses beyond the external aldimine 1 stage, the AFPA moiety loses the fluorine group, resulting in a reactive intermediate. Eventually, AFPA inactivates the PLP-dependent aminotransferase via a ternary adduct, in which the AFPA moiety is covalently linked to both PLP and the catalytic lysine, or alternatively forms external aldimine 3. The catalytic lysine freed after the formation of the external aldimine 1 plays a critical role in assisting the irreversible inactivation of the PLP enzyme. The formation of the proposed ternary adduct should provide evidence that the reaction between AFPA and PLP can progress beyond the external aldimine 1, reaching quinonoid or ketimine intermediates (Schemes 2 and Figure 29)



Scheme 2. Proposed Reaction Mechanism of AFPA. The green dashed box (1) shows the final adduct in the GabR Eb/O domain; the red dashed box (2) shows the expected adduct for Asp-AT; the blue dashed box (3) indicates an alternative external aldimine species

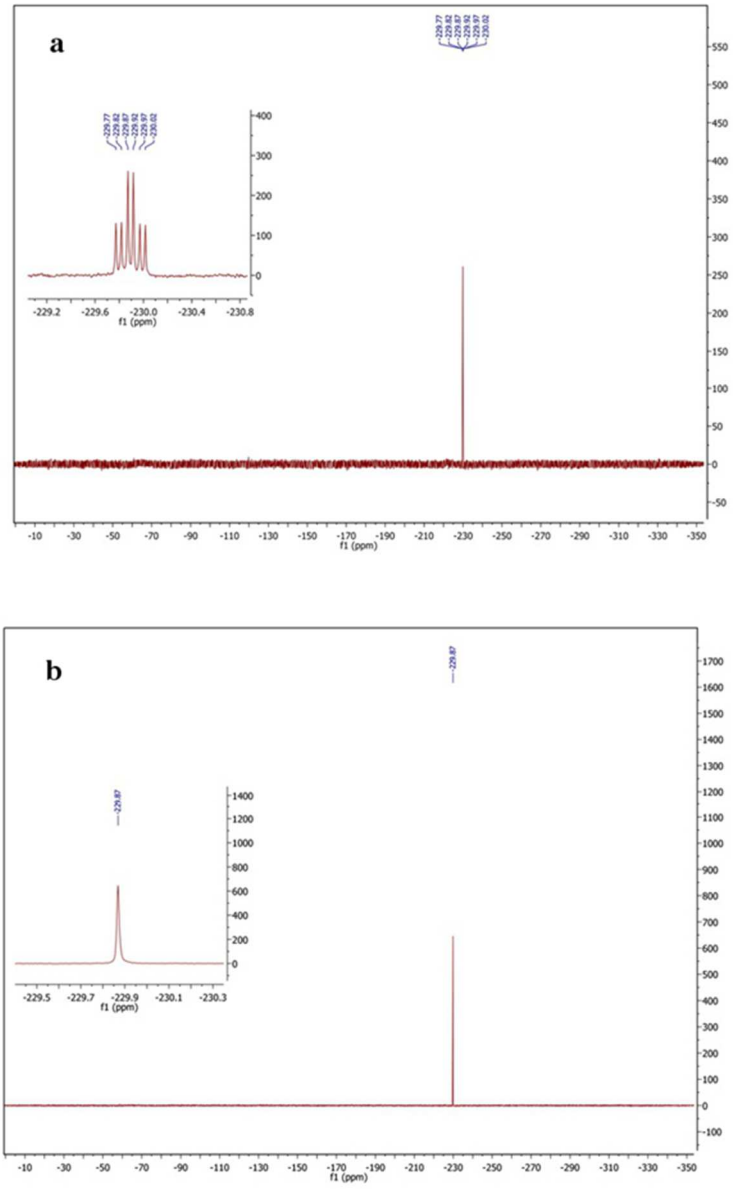


Figure 29. Fluorine NMR Spectra. a) Fluorine NMR showing a triplet of doublets (J_1 = 25.3 Hz, J_2 = 21.4 Hz). b) Proton-decoupled fluorine NMR spectrometry that shows a singlet.

In contrast, the intacted fluorine group of AFPA should indicate that the reaction cannot proceed beyond the formation of the external aldimine 1. *E coli* Asp-AT, a

homolog of the GabR Eb/O domain, has been used successfully as a model enzyme to study mechanism-based inactivators of PLP-dependent enzymes [54-56]. Due to the desirable crystallographic behavior of Asp-AT in various co-crystallization conditions, we chose it as a control for the co-crystallization experiment with AFPA.

Both the GabR Eb/O domain and Asp-AT were co-crystallized with PLP and AFPA. The structure of the Eb/O-PLP-AFPA complex (refined at 2.25 Å resolution) confirms the formation of external aldimine with the fluorine atom intact in the final adduct (Figure 30) To validate the existence of the fluorine atom, extra rounds of refinement were done after deletion of the fluorine atom; the omit map difference density (*fo-fc*) strongly supports the presence of the fluorine atom in the complex. Residue Y281 is observed to be in the same position as that in Eb/O-PLP-GABA complex structure (Figure 27), preventing the access of K312 to the PLP-AFPA Schiff base, which otherwise could potentially push the reaction further along the proposed inactivation pathway (Scheme 2).

Furthermore, the AFPA-treated protein sample was subjected to NMR studies to eliminate the possibility of an alternative adduct (adduct 3 in Scheme 2, via pathway b in Scheme 2). Pathway b and adduct 3 were never previously proven in the studies of the AFPA inactivation. However, it is chemically plausible that a hydroxyl or water could react with the reactive intermediate and replace the eliminated fluorine group; the resulting adduct 3 would be structurally similar to the external aldimine proposed here (adduct 1 in Scheme 2). At 2.25 Å resolution obtained for the complex structure Eb/O-

PLP-AFPA, it is impossible to distinguish adduct 3 from adduct 1 even with the omit map difference density (*fo-fc*) validation shown (as green mesh) in Figure 30a. Fluorine NMR spectrometry was hence employed to eliminate this hypothetical possibility. The results are shown in Figure 29. Figure S29a is the fluorine NMR spectrum that shows a triplet of doublets (J_1 = 25.3 Hz, J_2 = 21.4 Hz), since the fluorine atom is split by its two neighboring hydrogen atoms, then split again by another hydrogen atom on the adjacent carbon in the external aldimine form. Figure S1b represents the proton-decoupled fluorine NMR spectrum that shows a singlet, because the fluorine atom and the protons are decoupled. As concluded, the fluorine atom remains intact in the same environment as intact AFPA, even after treatment with a saturated amount of GabR protein.

As a positive control, the structure of the Asp-AT-PLP-AFPA complex (refined at 1.53 Å resolution) showed that the proposed ternary adduct 3 (has formed via pathway a in Scheme 2 (Figure 30b), indicating that the catalytic actions of the conserved lysine have occurred in Asp-AT but not in the GabR Eb/O domain. On the basis of structural information and NMR results, it was concluded that the external aldimine 1 is the final chemical species when both PLP and AFPA are bound to the GabR Eb/O domain. This suggests that the observed Schiff base in the Eb/O-PLP-GABA complex structure is also an external aldimine.

For aminotransferases in general, the external aldimine intermediate can be transiently observed spectroscopically [57]. However, mutation of a catalytic residue [58] or chemical modification of the ligands such as the amino acid [52] or PLP [59] was required to

trap a stable external aldimine in an aminotransferase. GabR is, to our knowledge, the first reported protein capable of forming a stable external aldimine between PLP and GABA for its function in transcription regulation. Only mild global conformational changes are introduced in the PLP-bound Eb/O domain upon binding to GABA and formation of the external aldimine. A least squares structural comparison revealed that the Eb/O domains of the GabR-PLP dimer (PDB: 4N0B) aligned well with those of the Eb/O-PLP-GABA complex or the Eb/O-PLP-AFPA complex, with RMSD values for Cα atoms around 1.12 Å and 1.13 Å, respectively. Conformational changes could be more pronounced in the linker region between Eb/O and wHTH domains, and within the wHTH domains in the full-length dimeric GabR; these conformational changes likely contribute to the triggering of transcription activation.

Functional Significance of External Aldimine Formation

To extrapolate the functional interpretation from the GabR Eb/O domain to the full length GabR, both *in vivo* and *in vitro* assays were conducted to compare the biological effects of GABA and AFPA. When examining GabR-mediated activation *in vivo* using a gabT-lacZ fusion [25](Figure 31a), both GABA and AFPA triggered GabR-mediated transcription activation in *B. subtilis* cells *In vitro*, GabR, PLP, RNA polymerase, and the gabT promoter-containing DNA fragments were mixed with radioactively labeled nucleotides as described previously [24]. The gabT transcript was formed upon addition of either GABA or AFPA (Figure 31b). In contrast to *in vivo* results, AFPA triggered the GabR-mediated transcription activation *in vitro* with a higher

potency than did GABA. The GabR-PLP-AFPA complex may be a better transcriptional activator than the GabR-PLP-GABA complex. However, the bacterial cellular uptake of GABA may be more efficient than that of AFPA, resulting in a more pronounced biological effect *in vivo* upon the addition of GABA.

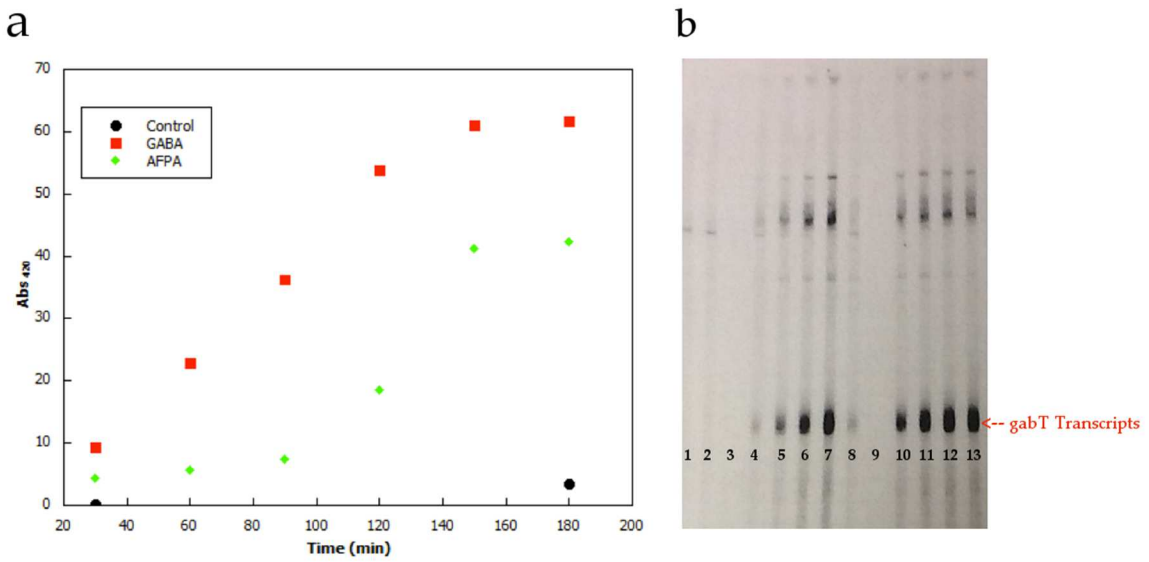


Figure 30. Biological effects of GABA and AFPA. a) The gabT-lacZ expression is monitored by absorbance at 420nM produced by the lacZ reaction using ONPG over time. b) Self-developed film indicating the newly synthesized gabT mRNA. Lane 1: Control without GABA or AFPA. Lanes 2-7: GABA added at concentration of 0.625, 1.25, 2.5, 5, 10 & 20 mM, respectively. Lanes 8-13: AFPA added with concentration of 0.625, 1.25, 2.5, 5, 10 & 20 mM, respectively.

Conclusions

We have shown that PLP bound to the GabR Eb/O domain reacts with the effector GABA to form a stable external aldimine, which apparently triggers the transcription from the gabT promoter in *B. subtilis*. This conclusion offers an explanation for the previous results that the effector pair of PMP and SSA can also trigger GabR-mediated transcription activation *in vitro* [24]. Through a reverse transamination-like reaction

(Scheme 1), SSA and PMP can potentially form a Schiff base, presenting a structure similar to the PLP-GABA external aldimine [24]. However, SSA and PMP may not work as an actual effector pair *in vivo*. With an evolutionarily modified PLP/GABA binding site, the GabR Eb/O domain is capable of preferentially binding to GABA and forming an external aldimine. To perform the role of a transcription regulator, GabR employs Y281 to subdue the Eb/O domain's catalytic potential by stabilizing the external aldimine formed and preventing its further modification.

In summary, these studies were submitted to the Proceedings of the National Academy of Sciences of the United States of America (PNAS) for publication: Rui Wu, Ruslan Sanishvili, Boris R. Belitsky, Jose I. Juncosa, Hoang V. Le, Helaina J.S. Lehrer, Michael Farley, Richard B. Silverman, Gregory A. Petsko, Dagmar Ringe, and Dali Liu. (2015)

Materials and Methods

Materials. Chemicals for the assays were purchased from Fisher Scientific (Pittsburgh, PA) and Sigma-Aldrich (St. Louis, MO); Cloning vectors were obtained from Integrated DNA Technologies (Coralville, IA). *Escherichia coli* BL21 (DE3) cells was purchased from New England BioLabs (Ipswich, MA). Bacterial growth media and antibiotics were obtained from Fisher Scientific and Sigma-Aldrich. Crystallization screen solutions and other crystallization supplies were purchased from Hampton Research (Aliso Viejo, CA) and Emerald Bio (Bedford, MA). All chemicals were of the highest quality available.

Truncation & Cloning of GabR Eb/O Domain. The sequence coding for GabR-Eb/O domain was amplified by PCR using previously cloned full-length GabR (in pETite vector from Lucigen) as the template [46], the specific oligonucleotides as primers (forward primer designed as GAAGGAGATATACATATGATCGACCAGAGCGATTGGATA and the reverse primer designed as GTGATGGTGGTGATGATGATCCCCTGTAACGGGGATTTT), and Phusion HF DNA polymerase, following the manufacture's instructions. The thermocycler program used included an initial denaturation for 120 s at 95 °C; 35 cycles of 95 °C for 30 s, 50 °C for 60 s, and 72 °C for 60 s; and a final extension of 72 °C for 5 min. The PCR product was inserted into a pETite vector (Lucigen) for recombinant expression of GabR Eb/O domain protein with a C-terminal His_6-tag.

Protein Expression and Purification. Full-length GabR protein and *L*-aspartate aminotransferase from *E. coli* were expressed and purified as described [24]. The GabR Eb/O domain was overexpressed in *E. coli* and purified using a similar method. Transformed E. coli BL21 (DE3) cells were grown in 1L of LB supplemented with 50 μg/ml kanamycin at 37 °C and 250 rpm until OD_{600} nm reached ~ 0.6. Protein expression was induced by the addition of 0.5 mM isopropyl-beta-D-1-thiogalactopyranoside (IPTG). Cells were incubated at 25 °C and harvested after 16 h by centrifuging at 4800 rpm and 4 °C for 10 min. The cell paste was resuspended in the lysis buffer containing 1 x PBS and 10 mM imidazole, pH 8.0 followed by sonication. The resulting suspension was centrifuged twice at 16,500 rpm and 4 °C for 20 min, and the soluble fraction was

loaded onto a 10 ml His-Trap column (GE Life Sciences, Piscataway, NJ) containing Ni^{2+} and previously equilibrated with lysis buffer. Elution of the retained proteins was achieved with a linear imidazole gradient (25 column volume, 10 mM-485 mM). Fractions containing truncated EB/O domain were pooled, concentrated to 2 mL, and loaded onto a 16/60 Superdex 200 size exclusion column (GE Life Sciences) that was previously equilibrated with 84% elution buffer. Fractions containing enzyme were pooled, concentrated, and used freshly after exchanging into suitable buffers for the crystallization and biological assays. The wild type Asp-AT was overexpressed and purified with the protocol used by Onuffer and Kirsch [60].

Crystallization of the GabR-Eb/O and X-ray Diffraction Data Collection. After the initial crystallization screen and optimization, the GabR-Eb/O domain was crystallized via hanging drop methods at 25 °C. The hanging drops were prepared with 1 μL of 12 mg/mL pre-incubated protein with inhibitor solution and 1 μL of the reservoir solution that contained 0.1 M Tris (pH 8.5) buffer and 20 % ethanol. Crystals were harvested in about 10 days. For cryo-cooling, crystals were transferred to the reservoir solution with 20 % glycerol as cryo protectant, and were flash-cooled by plunging in liquid nitrogen. The cocrystallization of Asp-AT and AFPA were conducted according to previous studies [54, 55] The diffraction data were collected at GM/CA@APS beamlines 23ID-B and 23ID-D at Advance Photon Source (Argonne National Laboratory, Argonne, IL). The wavelength used for data collection was 1.0332 Å, and diffraction images were

recorded on 4x4 tiled CCD detectors (Rayonix, Evanston, IL) with 300×300 mm^2 sensitive are. All data were indexed, integrated, scaled and merged using HKL2000 [41].

Phasing, Model Building, and Refinement. Diffraction data were phased with molecular replacement using the program Phaser [61]in the CCP4 software suite [42]. Previously solved full-length GabR structure [46], truncated to contain residues 107-479, was used as the search model. Rigid body refinement followed by restrained refinement was carried out in Refmac5. Model building was conducted using the program COOT [45]. Iterative rounds of model building and refinement generated the final coordinates.

In Vivo and in Vitro Assays for GabR-Dependent Transcription Activation. The *in vivo* assay was conducted as previously described [24], using GABA or AFPA as effectors (0.625 – 20 mM). *In vitro* GabR-dependent RNA polymerase reactions were performed as previously described [24]. The PCR-fragment (50nM) contained the gabR-gabT regulatory region. The total reaction was carried in transcription buffer containing: 40 mM Tris-HCl, 100 mM KCl, 10 mM $MgCl_2$, 5% glycerol, 0.1 mM EDTA, 1 mM DTT, 0,1 mg/mL of BSA with four units of RNasin, 150 μM ATP, CTP, and GTP, 20 μM UTP, 0.5 to 1 μl Ci of ^{32}P-labeled UTP, and 0.02-0.04 unit of *E. coli* RNA polymerase. The reaction was incubated at 37 ^{o}C for 15 min and terminated by additional of 4 μL of the 20 mM EDTA, 95% formamide dye solution. The samples were subsequently heated at 80 ^{o}C for 5 min. The samples were then analyzed without further purification using 7 M urea / 5-6% polyacrylamide DNA sequencing gels. GABA was purchased from Sigma. AFPA was synthesized according to previous publication [53].

Nuclear Magnetic Resonance (NMR) Spectroscopy. The protein (100 mg) was treated with AFPA. The molar ratio of the protein and AFPA was 2:1 to minimize free AFPA in the solution. The sample was concentrated to a volume of 0.4 mL by centrifugation. Deuterium oxide (0.15 mL) was added, and the solution was transferred to a 5 mm standard NMR tube. Fluorine NMR and proton-decoupled fluorine NMR spectra were taken at the Northwestern IMSERC facility on an Agilent DD2 500 MHz spectrometer with an Agilent 5 mm HFX probe at 26 ºC. ^{19}F NMR δ 229.87 (td, J_1=25.3 Hz, J_2=21.4 Hz); proton-decoupled ^{19}F NMR δ 229.87 (s).

Structural Analysis and Figure Making. All structural analyses were conducted in COOT [45] and UCSF Chimera [53]. Structure comparison of the PLP/amino acid binding site was conducted in two steps (Figure 2b & Figure 6): First, the overall structural comparison/overlap was conducted in UCSF Chimera using the Needleman-Wunsch algorithm. Second, manual adjustments were done by overlapping the pyridine ring of the PLP to achieve better pattern matching for the conserved key residues in the PLP/amino acid binding site for comparison. All structural figures were made in UCSF Chimera.

CHAPTER FOUR

CONCLUSIONS AND FUTURE INVESTIGATIONS

Full length GabR and Its Overall Structure

In conclusion, the studies presented herein have as provided useful information to elucidate the transcription regulatorty mechanism of GabR, gain insight into the functionality of a unique MocR subfamily member, gain insight into the regulation pathway of the GABA shunt pathway, and provided some detailed molecular and chemical information for future anti-vurilance drug discovery.

The obtained X-ray crystal structures of full length GabR provided information on the domain arragements of MocR subfamily's transcriotion. To our best knowledge, these structures are the first X-ray crystal structures revealed in the GntR/MocR family. In addition, we also bead model obtained from the small angle X-ray scattering (SAXS) from protein solution sample (Figure 31), and structure impose indicated a high agreement between the bead model and the X-ray crystal structure (Figure 32). These suggested that the head-to-tail homodimer is probably physciology and in the functional geometry before transcription initiation. More detailed structure analysis provides offer more information:

(1) To allow us to propose a model for the DNA and protein complex, further to spectulated how the interaction between DNA and GabR could position the regulatory domain in the proper position and set the "perfect" geometry to initiate the transcription.

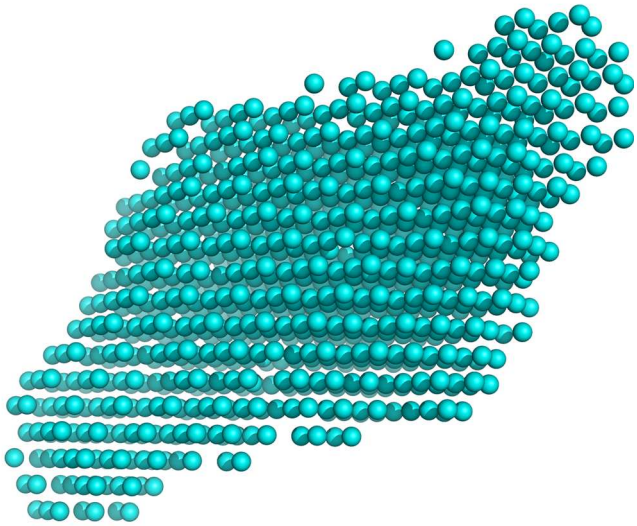


Figure 31. Small Angle X-ray Scattering (SAXS) bead model of full length GabR.

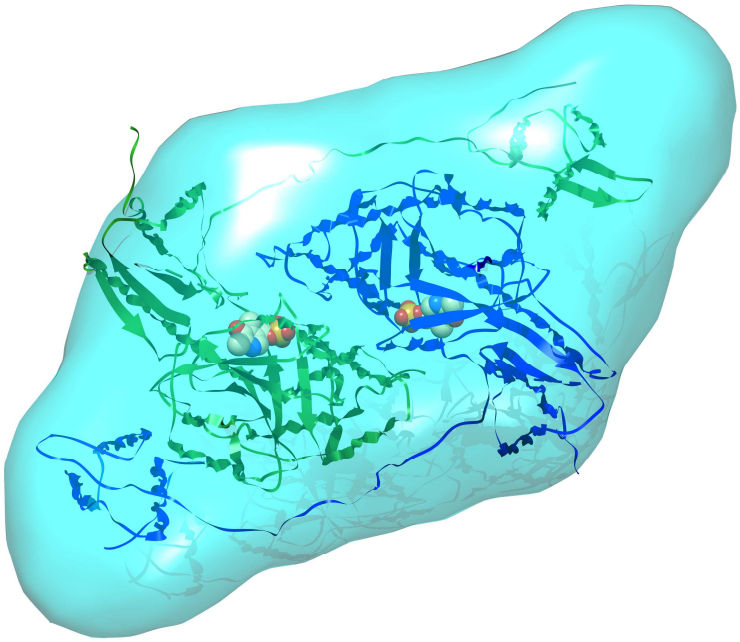


Figure 32. Small Angle X-ray Scattering (SAXS) bead model of full length GabR superimpose with X-ray crystal structure of full length GabR.

(2) The reveal of the effector/s (GABA and PLP) binding pocket in effector binding and/or regulatory domain offer us a bold hypothesis that nature may recycled the vitamin B6 and "borrow" a partial reaction from aminotransferase to regulate the

\

transcription initation.

(3) The linker region between the DNA binding domain and the effectors binding and /or regulatory domain shown to be the first case in GntR transcription regulator super family. The long linker suggestes a rather dramatic confirmation changes could happen upon binding to the effectos during the initiation of the transcription.

Effector Binding Domain and Related Biological Function

Although these structures have provided much needed answers in terms of static image before GabR associate with the effector GABA. There are still more details that need to be addressed to elucidate how effector triggers the transcription initiation.

Further studies reveavled several Eb/O domain structures in complex with GABA, and molecular probe molecule AFPA. These two structures paved the way toward understanding how GabR functions. Our results suggest that upon binding to effector (GABA) or effectors analog (APFA), the internal aldimine would undergo a reaction with coming effector to form external aldimine. Using multiple approaches, such as: NMR, *in vivo* transcripton assay, *in vitro* transcription assay, we were able to determine that the external aldimine is a critical chemical species that triggers the transcription initation.

GabR and DNA Complex

One of the disadvantage of having only the Eb/O domain is no structural information could support during the initiation of the transcription if there is a conformational changes on the geometry between Eb/O and DNA binding domain, or if

the DNA binding domain have re-located upon binding to effectors. This will hopefully be addressed in the near future. Meanwhile a SAXS bead modeal of the GabR/DNA complex (Figure 33 and Figure 34) shed some light on how the DNA and GabR would interact. Noteably, the DNA fragment is at the bented confirmation from the interpretation from the SAXS bead model. Looks apparent to us this bented geometry somehow facilitate the transcription initiation, more specifically maybe facilitating RNA polymerase, sigma factor and other important component to bind to initiate the transcription.

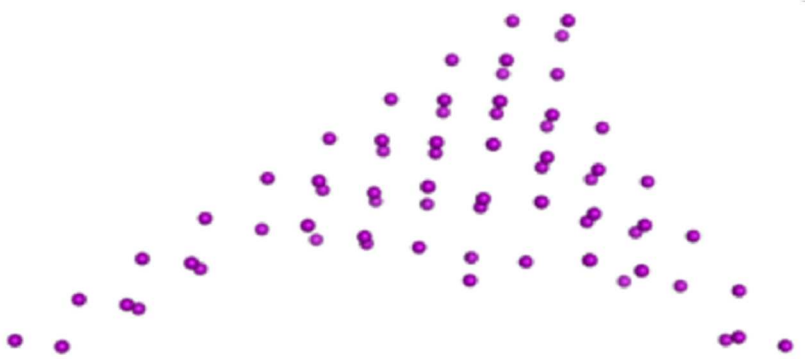


Figure 33. Small Angle X-ray Scattering (SAXS) bead model of full length GabR and DNA complex.

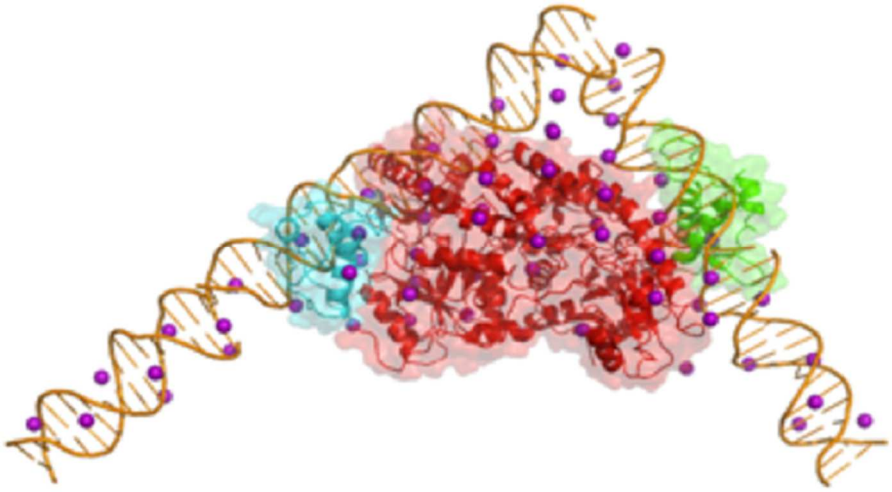


Figure 34. The fittng model of Small Angle X-ray Scattering (SAXS) bead stcuture of full length GabR and DNA complex and X-ray crystal structure of GabR.

CHAPTER FIVE

THE CRYSTAL STRUCTURE OF NITROSOMONAS EUROPAEA SUCROSE SYNTHASE REVEALS CRITICAL CONFORMATIONAL CHANGES AND INSIGHTS INTO THE SUCROSE METABOLISM IN PROKARYOTES

Research Significance

In this chapter we report the first crystal structure of a prokaryotic sucrose synthase from the non-photosynthetic bacterium *Nitrosomonas europaea*. The obtained structure was in an open form, whereas the only other available structure from the plant *Arabidopsis thaliana* was in a closed conformation. Comparative structural analysis revealed a "hinge-latch" combination, which is critical to transition between the open and closed forms of the enzyme. The *N. europaea* sucrose synthase shares the same fold as the GT-B family of the retaining glycosyltransferases. In addition, a triad of conserved homologous catalytic residues in the family showed to be functionally critical in the *N. europaea* sucrose synthase (Arg567, Lys572, Glu663). This implies that sucrose synthase shares not only a common origin with the GT-B family, but also a similar catalytic mechanism. The enzyme was proven to prefer transferring glucose from ADP-glucose rather than UDP-glucose like the eukaryotic counterparts. This predicts that these prokaryotic organisms have a different sucrose metabolic scenario from plants. Nucleotide preference determines where the glucose moiety is targeted after sucrose is degraded.

Introduction

In plants, sucrose (Suc) is a major photosynthetic product and plays a key role not only for carbon partition but also in sugar sensing, development, and regulation of gene expression [62-64]. It was first thought that Suc metabolism was a characteristic of plants but it was later found in other oxygenic photosynthetic organisms [65, 66]. In the last decade, Salerno and coworkers demonstrated the importance of Suc for carbon and nitrogen fixation in filamentous cyanobacteria [67, 68]. More recently, genomic and phylogenetic analyses revealed the existence of Suc-related genes in non-photosynthetic prokaryotes such as proteobacteria, firmicutes, and planctomycetes [65, 66, 69]. It has been suggested that these organisms acquired the genes of Suc metabolism by horizontal gene transfer [65, 66, 69]. However, analysis of the enzymes encoded by such genes is currently lacking.

Nitrosomonas europaea is a chemo-litho-autotrophic bacterium that obtains energy by oxidizing ammonia to hydroxylamine and nitrite in presence of oxygen [70]. It is a member of the β-proteobacteria group with a putative photosynthetic ancestor [71]. *N. europaea* has potential for many biotechnological applications, including bioremediation of water contaminated with chlorinated aliphatic hydrocarbons [72] or ammonia, in combination with *Paracoccus denitrifi* [70]. *N. europaea* displays some metabolic resemblance to photosynthetic organisms, but with marked differences. For instance, it possesses all the coding genes for enzymes of the Calvin-Benson cycle, but with two exceptions that could be replaced by other glycolytic enzymes [73]. All the genes coding for enzymes from the tricarboxylic acid cycle were found in *N. europaea* [73]; however, activity of α-ketoglutarate dehydrogenase is non-detectable [74].

The evidence from genomic studies suggests that *N. europaea* can synthesize Suc [73]; however, the biochemical properties of enzymes from Suc metabolism have not been characterized. Generally, in plants, Suc is synthesized from UDP-glucose (UDP-Glc) and fructose-6-phosphate (Fru-6P) in a reaction catalyzed by Suc-6-phosphate synthase (EC 2.4.1.14), followed by removal of the phosphate group by Suc-6-phosphatase (EC 3.1.3.24). The disaccharide can be degraded to Glc and Fru by invertases (EC 3.2.1.26) or cleaved by UDP to form UDP-Glc and Fru by Suc synthase (SucSase, EC 2.4.1.13) [63, 64]. However, some plant SucSases have a certain degree of substrate promiscuity [75-82], and *Thermosynechococcus elongatus* prefers ADP [77]. For that reason, a general reversible reaction could be written as:

$$NDP + Suc \rightleftharpoons NDP\text{-}Glc + Fru$$

Besides its physiological role, SucSase catalyzes a reversible reaction and its activity can be measured in both directions *in vitro*. In filamentous cyanobacteria, the products derived from Suc cleavage contribute to other biological processes, such as polysaccharides synthesis [83]. Therefore, understanding the catalysis and the regulation of SucSase is of great significance. Recently, Zheng et al. [84] reported the crystal structure of the *Arabidopsis thaliana* SucSase (*Ath*SucSase) in complex with UDP and fructose in a closed conformation. This enzyme is a homotetramer composed of four identical subunits of ~90 kDa and belongs to group 4 of the GT-B retaining glycosyltransferase family (http://www.cazy.org/GlycosylTransferases.html) [85]. The reaction mechanism for this enzyme family has not been elucidated. But, it has been suggested that it favors the S_Ni-like reaction mechanism proposed by Lairson et al[84-86].

Although several cyanobacterial [69, 77, 80] and plant [75, 78, 87-90] SucSases have been characterized, the enzyme from non-photosynthetic bacteria has never been studied and no structural information of any SucSase from bacterial sources is available. In this work we report the recombinant expression and biochemical characterization of *N. europaea* SucSase (*Neu*SucSase) and its crystal structure. We also determined the catalytic implications of highly conserved residues and the specificity for nucleotide substrates.

Sequence Analysis

To know how *Neu*SucSase relates to other SucSases from divergent organisms we constructed a phylogenetic tree using 117 amino acid sequences retrieved from the NCBI database (Figure 35, Table 3, and Figure 37). The tree comprised seven major branches, containing the sequences from cyanobacteria (group I; 21 sequences), proteobacteria (groups II and III; 17 sequences), the moss *Physcomitrella patens* subsp. patens (group IV; 4 sequences), and vascular plants (groups V, VI, and VII; 75 sequences) (Figure 35). The shape of the tree shown in Figure 35 is similar to the one published by Kolman et al. [69]. Group I is subdivided in two branches, containing the sequences encoded by the *susA* (cyan) and *susB* (orange) genes (Figure 35) [69]. Most sequences from proteobacteria are included in group III (including β-, γ- and δ-)

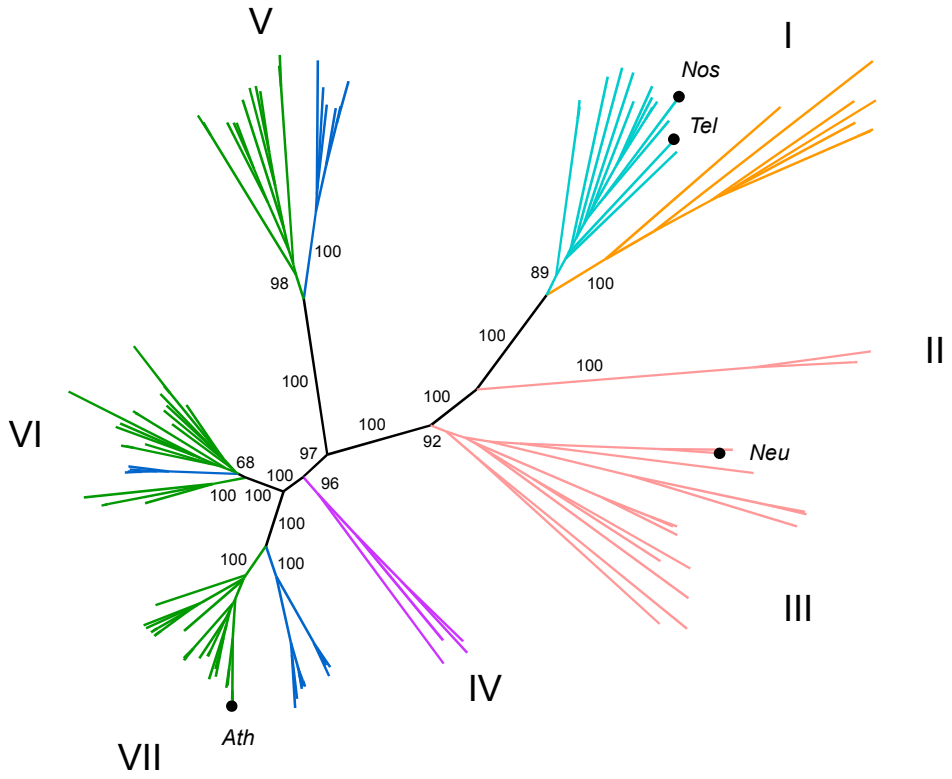


Figure 35. Phylogenetic tree of SucSases. Group I contains sequences from cyanobacteria and is divided in two major branches, susA (cyan) and susB (orange) proteins; groups II and III (pink) contain sequences from proteobacteria; group IV contains the sequences from the moss *P. patens* (violet); and groups V, VI, and VII, which contain the sequences from vascular plants are further divided in two branches, dicots (green) and monocots (blue). Numbers for major branches are the bootstrap values obtained during tree reconstruction, as described in the "Materials and Methods" section. *Neu*, *N. europaea*; *Ath*, *A. thaliana* SucSase 1; *Nos*, *Nostoc* sp. PCC 7120 susA; *Tel*, *T. elongatus*. proteobacteria); though, the sequences from δ-proteobacterium MLMS-1 and *Desulfurivibrio alkaliphilus* AHT2 are in a diverging branch (group II)

Table 3. Sucrose synthase genes used for phylogenetic analysis. Different colors belong to different branches in Fig. 35.

Accession (GI)	Organism	Phylum	Type	Annotation
158339122	*Acaryochloris marina* MBIC11017	cyanobacteria		sucrose synthase
75908500	*Anabaena variabilis* ATCC 29413	cyanobacteria		sucrose synthase
209523126	*Arthrospira maxima* CS-328	cyanobacteria		sucrose synthase
284053161	*Arthrospira platensis* str. Paraca	cyanobacteria		sucrose synthase
218440696	*Cyanothece* sp. PCC 7424	cyanobacteria		sucrose synthase
220907171	*Cyanothece* sp. PCC 7425	cyanobacteria		sucrose synthase
332712456	*Lyngbya majuscula* 3L	cyanobacteria		sucrose synthase
254416162	*Microcoleus chthonoplastes* PCC 7420	cyanobacteria		sucrose synthase
334117431	*Microcoleus vaginatus* FGP-2	cyanobacteria		sucrose synthase
159031025	*Microcystis aeruginosa* PCC 7806	cyanobacteria		unnamed protein product
119512682	*Nodularia spumigena* CCY9414	cyanobacteria		sucrose synthase
186682280	*Nostoc punctiforme* PCC 73102	cyanobacteria		sucrose synthase
17232477	*Nostoc* sp. PCC 7120	cyanobacteria		sucrose synthase
22298591	*Thermosynechococcus elongatus* BP-1	cyanobacteria		sucrose synthase
75909957	*Anabaena variabilis* ATCC 29413	cyanobacteria		sucrose synthase
220909283	*Cyanothece* sp. PCC 7425	cyanobacteria		sucrose synthase
332708740	*Lyngbya majuscula* 3L	cyanobacteria		sucrose synthase
119510624	*Nodularia spumigena* CCY9414	cyanobacteria		sucrose synthase
298489784	*Nostoc azollae* 0708	cyanobacteria		sucrose synthase
186685043	*Nostoc punctiforme* PCC 73102	cyanobacteria		sucrose synthase

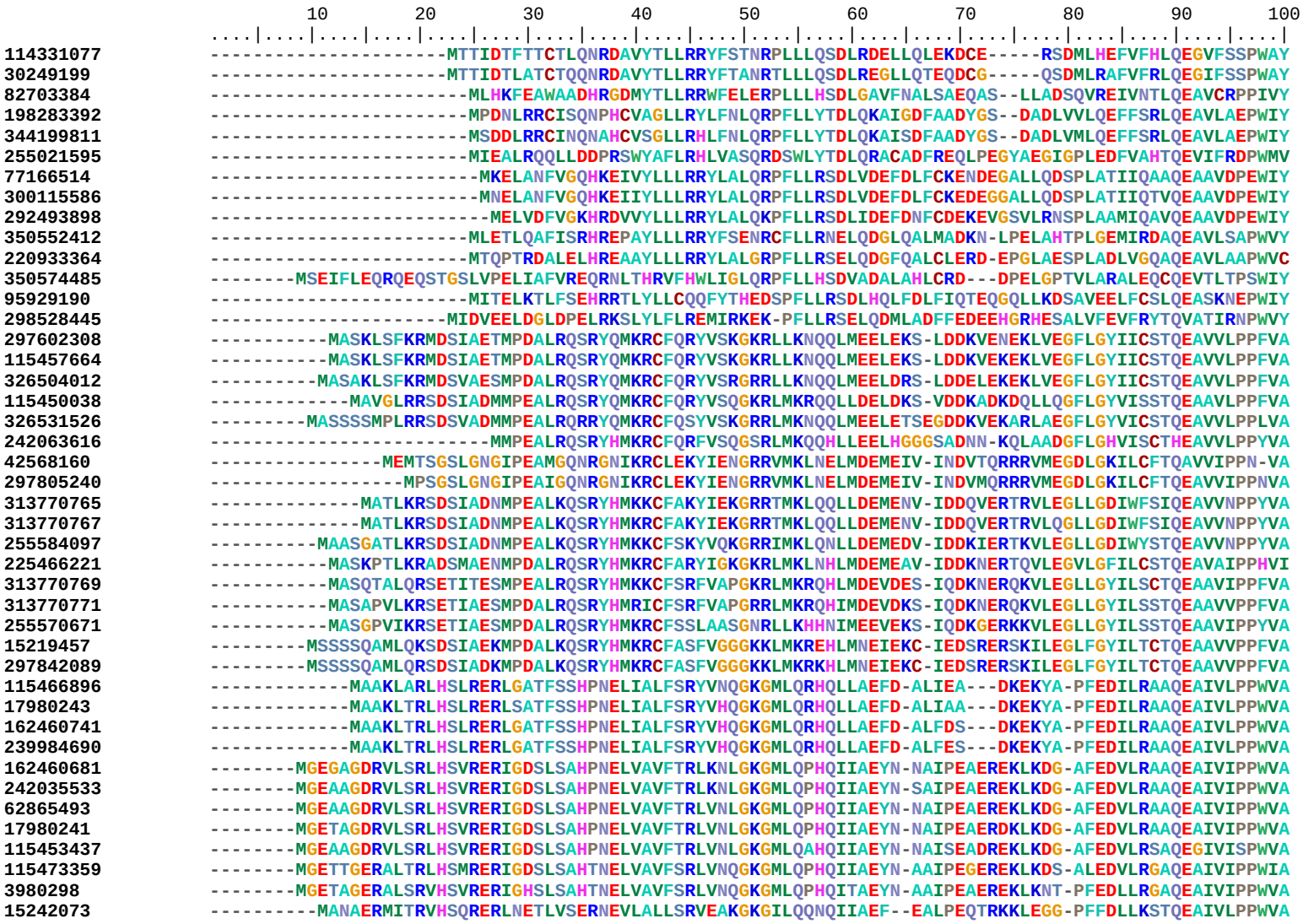
17228554	*Nostoc* sp. PCC 7120	cyanobacteria		sucrose synthase
168009716	*Physcomitrella patens* subsp. patens	streptophyta	moss	predicted protein
168029793	*Physcomitrella patens* subsp. patens	streptophyta	moss	predicted protein
168035060	*Physcomitrella patens* subsp. patens	streptophyta	moss	predicted protein
168058907	*Physcomitrella patens* subsp. patens	streptophyta	moss	predicted protein
297795665	*Arabidopsis lyrata* subsp. lyrata	streptophyta	eudicot	hypothetical protein
297805240	*Arabidopsis lyrata* subsp. lyrata	streptophyta	eudicot	hypothetical protein
297812265	*Arabidopsis lyrata* subsp. lyrata	streptophyta	eudicot	hypothetical protein
297814081	*Arabidopsis lyrata* subsp. lyrata	streptophyta	eudicot	hypothetical protein
297818772	*Arabidopsis lyrata* subsp. lyrata	streptophyta	eudicot	hypothetical protein
297842089	*Arabidopsis lyrata* subsp. lyrata	streptophyta	eudicot	hypothetical protein
15219457	*Arabidopsis thaliana*	streptophyta	eudicot	sucrose synthase 6
42568160	*Arabidopsis thaliana*	streptophyta	eudicot	sucrose synthase 5
22331535	*Arabidopsis thaliana*	streptophyta	eudicot	sucrose synthase 4
15235300	*Arabidopsis thaliana*	streptophyta	eudicot	sucrose synthase 3
15239816	*Arabidopsis thaliana*	streptophyta	eudicot	sucrose synthase 2
15242073	*Arabidopsis thaliana*	streptophyta	eudicot	sucrose synthase 1
151176306	*Beta vulgaris*	streptophyta	eudicot	sucrose synthase 1
14530225	*Beta vulgaris*	streptophyta	eudicot	sucrose synthase
38425095	*Beta vulgaris*	streptophyta	eudicot	sucrose synthase
6682841	*Citrus unshiu*	streptophyta	eudicot	sucrose synthase
6683114	*Citrus unshiu*	streptophyta	eudicot	sucrose synthase
115310620	*Coffea arabica*	streptophyta	eudicot	sucrose synthase
115310618	*Coffea arabica*	streptophyta	eudicot	sucrose synthase

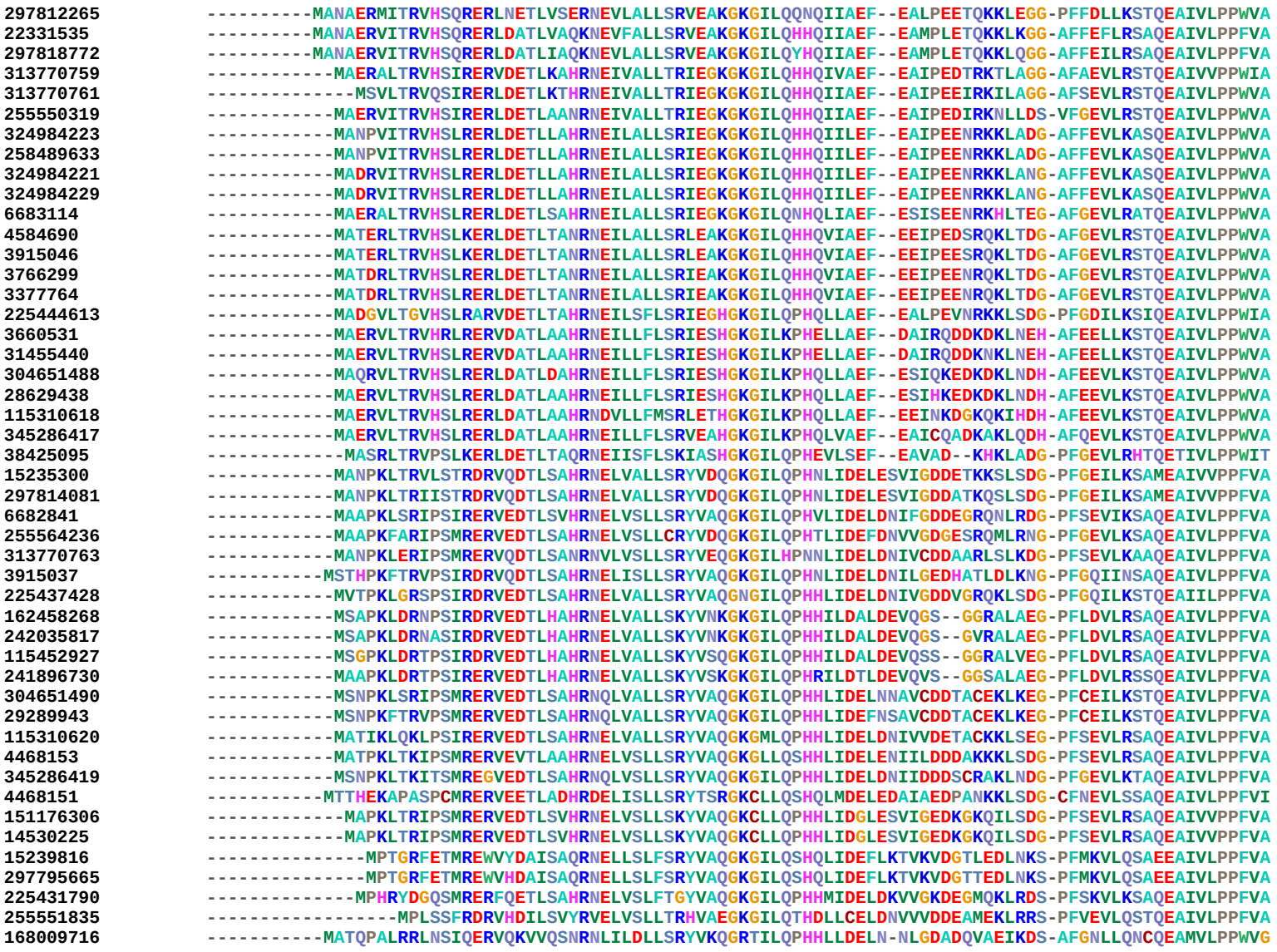
4468153	*Craterostigma plantagineum*	streptophyta	eudicot	sucrose synthase
4468151	*Craterostigma plantagineum*	streptophyta	eudicot	sucrose synthase
324984223	*Gossypium barbadense*	streptophyta	eudicot	sucrose synthase
324984221	*Gossypium barbadense*	streptophyta	eudicot	sucrose synthase
324984229	*Gossypium hirsutum*	streptophyta	eudicot	sucrose synthase
258489633	*Gossypium hirsutum*	streptophyta	eudicot	sucrose synthase 1
3915046	*Medicago sativa*	streptophyta	eudicot	sucrose synthase
4584690	*Medicago truncatula*	streptophyta	eudicot	sucrose synthase
345286419	*Orobanche ramosa*	streptophyta	eudicot	sucrose synthase 2
345286417	*Orobanche ramosa*	streptophyta	eudicot	sucrose synthase 1
3915037	*Pisum sativum*	streptophyta	eudicot	sucrose synthase 2
3766299	*Pisum sativum*	streptophyta	eudicot	sucrose synthase
3377764	*Pisum sativum*	streptophyta	eudicot	nodule-enhanced susy
313770771	*Populus trichocarpa*	streptophyta	eudicot	sucrose synthase 7
313770769	*Populus trichocarpa*	streptophyta	eudicot	sucrose synthase 6
313770767	*Populus trichocarpa*	streptophyta	eudicot	sucrose synthase 5
313770765	*Populus trichocarpa*	streptophyta	eudicot	sucrose synthase 4
313770763	*Populus trichocarpa*	streptophyta	eudicot	sucrose synthase 3
313770761	*Populus trichocarpa*	streptophyta	eudicot	sucrose synthase 2
313770759	*Populus trichocarpa*	streptophyta	eudicot	sucrose synthase 1
255550319	*Ricinus communis*	streptophyta	eudicot	sucrose synthase
255551835	*Ricinus communis*	streptophyta	eudicot	sucrose synthase
255564236	*Ricinus communis*	streptophyta	eudicot	sucrose synthase
255570671	*Ricinus communis*	streptophyta	eudicot	sucrose synthase
255584097	*Ricinus communis*	streptophyta	eudicot	sucrose synthase

304651490	*Solanum lycopersicum*	streptophyta	eudicot	sucrose synthase isoform 4
304651488	*Solanum lycopersicum*	streptophyta	eudicot	sucrose synthase isoform 3
3660531	*Solanum lycopersicum*	streptophyta	eudicot	sucrose synthase
31455440	*Solanum tuberosum*	streptophyta	eudicot	sucrose synthase 4
28629438	*Solanum tuberosum*	streptophyta	eudicot	sucrose synthase 2
29289943	*Solanum tuberosum*	streptophyta	eudicot	sucrose synthase
225431790	*Vitis vinifera*	streptophyta	eudicot	hypothetical protein
225437428	*Vitis vinifera*	streptophyta	eudicot	hypothetical protein
225444613	*Vitis vinifera*	streptophyta	eudicot	hypothetical protein
225466221	*Vitis vinifera*	streptophyta	eudicot	hypothetical protein
62865493	*Bambusa oldhamii*	streptophyta	monocot	sucrose synthase
17980241	*Bambusa oldhamii*	streptophyta	monocot	sucrose synthase
17980243	*Bambusa oldhamii*	streptophyta	monocot	sucrose synthase
326531526	*Hordeum vulgare* subsp. vulgare	streptophyta	monocot	predicted protein
326504012	*Hordeum vulgare* subsp. vulgare	streptophyta	monocot	predicted protein
241896730	*Hordeum vulgare* subsp. vulgare	streptophyta	monocot	sucrose synthase
3980298	*Hordeum vulgare* subsp. vulgare	streptophyta	monocot	sucrose synthase 2
239984690	*Hordeum vulgare* subsp. vulgare	streptophyta	monocot	sucrose synthase
115450038	*Oryza sativa* Japonica Group	streptophyta	monocot	Os02g0831500
115452927	*Oryza sativa* Japonica Group	streptophyta	monocot	Os03g0340500
115453437	*Oryza sativa* Japonica Group	streptophyta	monocot	Os03g0401300
115457664	*Oryza sativa* Japonica Group	streptophyta	monocot	Os04g0309600
115466896	*Oryza sativa* Japonica Group	streptophyta	monocot	Os06g0194900
115473359	*Oryza sativa* Japonica Group	streptophyta	monocot	Os07g0616800
297602308	*Oryza sativa* Japonica Group	streptophyta	monocot	Os04g0249500

242063616	*Sorghum bicolor*	streptophyta	monocot	hypothetical protein
242035533	*Sorghum bicolor*	streptophyta	monocot	hypothetical protein
242035817	*Sorghum bicolor*	streptophyta	monocot	hypothetical protein
162458268	*Zea mays*	streptophyta	monocot	sucrose synthase 2
162460681	*Zea mays*	streptophyta	monocot	sucrose synthase 2
162460741	*Zea mays*	streptophyta	monocot	sucrose synthase 1
114331077	*Nitrosomonas eutropha* C91	β-proteobacteria		sucrose synthase
82703384	*Nitrosospira multiformis* ATCC 25196	β-proteobacteria		sucrose synthase
30249199	*Nitrosomonas europaea* ATCC 19718	β-proteobacteria		sucrose synthase
94264333	delta proteobacterium MLMS-1	δ-proteobacteria		sucrose synthase
94266940	delta proteobacterium MLMS-1	δ-proteobacteria		sucrose synthase
95929190	*Desulfuromonas acetoxidans* DSM 684	δ-proteobacteria		sucrose synthase
297569306	*Desulfurivibrio alkaliphilus* AHT2	δ-proteobacteria		sucrose synthase
298528445	*Desulfonatronospira thiodismutans* ASO3-1	δ-proteobacteria		sucrose synthase
77166514	*Nitrosococcus oceani* ATCC 19707	γ-proteobacteria		sucrose synthase
292493898	*Nitrosococcus halophilus* Nc4	γ-proteobacteria		sucrose synthase
300115586	*Nitrosococcus watsonii* C-113	γ-proteobacteria		sucrose synthase
350552412	*Thiorhodospira sibirica* ATCC 700588	γ-proteobacteria		sucrose synthase
350574485	*Thiorhodovibrio* sp. 970	γ-proteobacteria		sucrose synthase
220933364	*Thioalkalivibrio sulfidophilus* HL-EbGr7	γ-proteobacteria		sucrose synthase
255021595	*Acidithiobacillus caldus* ATCC 51756	γ-proteobacteria		sucrose synthase
198283392	*Acidithiobacillus ferrooxidans* ATCC 53993	γ-proteobacteria		sucrose synthase

Figure 36. Alignment of sucrose synthases from Table 3.



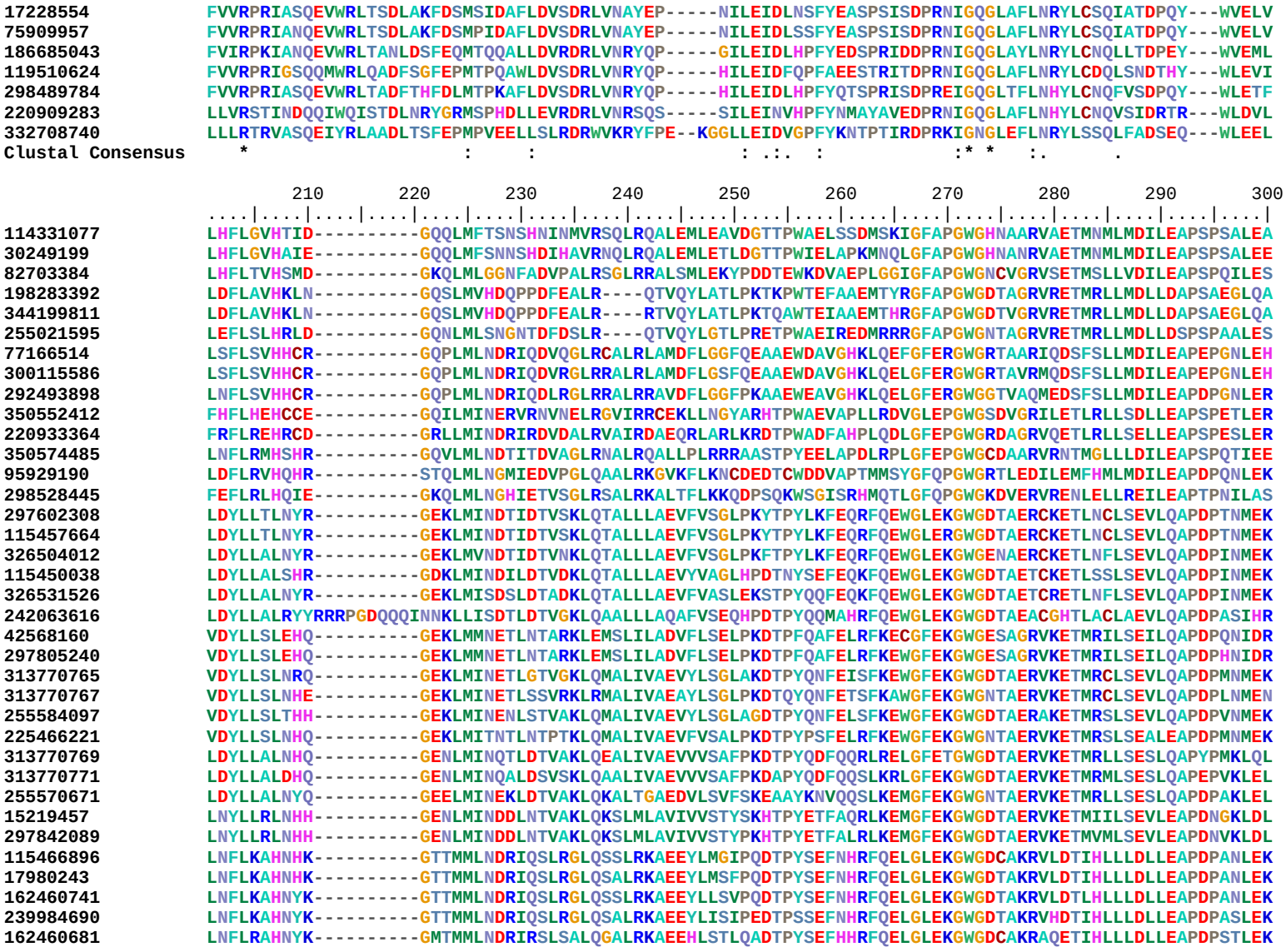

297812265 ----------MANAERMITRVHSQRERLNETLVSERNEVLALLSRVEAKGKGILQQNQIIAEF--EALPEETQKKLEGG-PFFDLLKSTQEAIVLPPWVA
22331535 ----------MANAERVITRVHSQRERLDATLVAQKNEVFALLSRVEAKGKGILQHHQIIAEF--EAMPLETQKKLKGG-AFFEFLRSAQEAIVLPPFVA
297818772 ----------MANAERVITRVHSQRERLDATLIAQKNEVLALLSRVEAKGKGILQYHQIIAEF--EAMPLETQKKLQGG-AFFEILRSAQEAIVLPPFVA
313770759 -----------MAERALTRVHSIRERVDETLKAHRNEIVALLTRIEGKGKGILQHHQIVAEF--EAIPEDTRKTLAGG-AFAEVLRSTQEAIVVPPWIA
313770761 --------------MSVLTRVQSIRERLDETLKTHRNEIVALLTRIEGKGKGILQHHQIIAEF--EAIPEEIRKILAGG-AFSEVLRSTQEAIVLPPWVA
255550319 -----------MAERVITRVHSIRERLDETLAANRNEIVALLTRIEGKGKGILQHHQIIAEF--EAIPEDIRKNLLDS-VFGEVLRSTQEAIVLPPWVA
324984223 -----------MANPVITRVHSLRERLDETLLAHRNEILALLSRIEGKGKGILQHHQIILEF--EAIPEENRKKLADG-AFFEVLKASQEAIVLPPWVA
258489633 -----------MANPVITRVHSLRERLDETLLAHRNEILALLSRIEGKGKGILQHHQIILEF--EAIPEENRKKLADG-AFFEVLKASQEAIVLPPWVA
324984221 -----------MADRVITRVHSLRERLDETLLAHRNEILALLSRIEGKGKGILQHHQIILEF--EAIPEENRKKLANG-AFFEVLKASQEAIVLPPWVA
324984229 -----------MADRVITRVHSLRERLDETLLAHRNEILALLSRIEGKGKGILQHHQIILEF--EAIPEENRKKLANG-AFFEVLKASQEAIVLPPWVA
6683114 -----------MAERALTRVHSLRERLDETLSAHRNEILALLSRIEGKGKGILQNHQLIAEF--ESISEENRKHLTEG-AFGEVLRATQEAIVLPPWVA
4584690 -----------MATERLTRVHSLKERLDETLTANRNEILALLSRLEAKGKGILQHHQVIAEF--EEIPEDSRQKLTDG-AFGEVLRSTQEAIVLPPWVA
3915046 -----------MATERLTRVHSLKERLDETLTANRNEILALLSRLEAKGKGILQHHQVIAEF--EEIPEESRQKLTDG-AFGEVLRSTQEAIVLPPWVA
3766299 -----------MATDRLTRVHSLRERLDETLTANRNEILALLSRIEAKGKGILQHHQVIAEF--EEIPEENRQKLTDG-AFGEVLRSTQEAIVLPPWVA
3377764 -----------MATDRLTRVHSLRERLDETLTANRNEILALLSRIEAKGKGILQHHQVIAEF--EEIPEENRQKLTDG-AFGEVLRSTQEAIVLPPWVA
225444613 -----------MADGVLTGVHSLRARVDETLTAHRNEILSFLSRIEGHGKGILQPHQLLAEF--EALPEVNRKKLSDG-PFGDILKSIQEAIVLPPWIA
3660531 -----------MAERVLTRVHRLRERVDATLAAHRNEILLFLSRIESHGKGILKPHELLAEF--DAIRQDDKDKLNEH-AFEELLKSTQEAIVLPPWVA
31455440 -----------MAERVLTRVHSLRERVDATLAAHRNEILLFLSRIESHGKGILKPHELLAEF--DAIRQDDKNKLNEH-AFEELLKSTQEAIVLPPWVA
304651488 -----------MAQRVLTRVHSLRERLDATLDAHRNEILLFLSRIESHGKGILKPHQLLAEF--ESIQKEDKDKLNDH-AFEEVLKSTQEAIVLPPWVA
28629438 -----------MAERVLTRVHSLRERLDATLAAHRNEILLFLSRIESHGKGILKPHQLLAEF--ESIHKEDKDKLNDH-AFEEVLKSTQEAIVLPPWVA
115310618 -----------MAERVLTRVHSLRERLDATLAAHRNDVLLFMSRLETHGKGILKPHQLLAEF--EEINKDGKQKIHDH-AFEEVLKSTQEAIVLPPWVA
345286417 -----------MAERVLTRVHSLRERLDATLAAHRNEILLFLSRVEAHGKGILKPHQLVAEF--EAICQADKAKLQDH-AFQEVLKSTQEAIVLPPWVA
38425095 -------------MASRLTRVPSLKERLDETLTAQRNEIISFLSKIASHGKGILQPHEVLSEF--EAVAD--KHKLADG-PFGEVLRHTQETIVLPPWIT
15235300 -----------MANPKLTRVLSTRDRVQDTLSAHRNELVALLSRYVDQGKGILQPHNLIDELESVIGDDETKKSLSDG-PFGEILKSAMEAIVVPPFVA
297814081 -----------MANPKLTRIISTRDRVQDTLSAHRNELVALLSRYVDQGKGILQPHNLIDELESVIGDDATKQSLSDG-PFGEILKSAMEAIVVPPFVA
6682841 -----------MAAPKLSRIPSIRERVEDTLSVHRNELVSLLSRYVAQGKGILQPHVLIDELDNIFGDDEGRQNLRDG-PFSEVIKSAQEAIVLPPFVA
255564236 -----------MAAPKFARIPSMRERVEDTLSAHRNELVSLLCRYVDQGKGILQPHTLIDEFDNVVGDGESRQMLRNG-PFGEVLKSAQEAIVLPPFVA
313770763 -----------MANPKLERIPSMRERVQDTLSANRNVLVSLLSRYVEQGKGILHPNNLIDELDNIVCDDAARLSLKDG-PFSEVLKAAQEAIVLPPFVA
3915037 -----------MSTHPKFTRVPSIRDRVQDTLSAHRNELISLLSRYVAQGKGILQPHNLIDELDNILGEDHATLDLKNG-PFGQIINSAQEAIVLPPFVA
225437428 -----------MVTPKLGRSPSIRDRVEDTLSAHRNELVALLSRYVAQGNGILQPHHLIDELDNIVGDDVGRQKLSDG-PFGQILKSTQEAIILPPFVA
162458268 -----------MSAPKLDRNPSIRDRVEDTLHAHRNELVALLSKYVNKGKGILQPHHILDALDEVQGS--GGRALAEG-PFLDVLRSAQEAIVLPPFVA
242035817 -----------MSAPKLDRNASIRDRVEDTLHAHRNELVALLSKYVNKGKGILQPHHILDALDEVQGS--GVRALAEG-PFLDVLRSAQEAIVLPPFVA
115452927 -----------MSGPKLDRTPSIRDRVEDTLHAHRNELVALLSKYVSQGKGILQPHHILDALDEVQSS--GGRALVEG-PFLDVLRSAQEAIVLPPFVA
241896730 -----------MAAPKLDRTPSIRERVEDTLHAHRNELVALLSKYVSKGKGILQPHRILDTLDEVQVS--GGSALAEG-PFLDVLRSSQEAIVLPPFVA
304651490 -----------MSNPKLSRIPSMRERVEDTLSAHRNQLVALLSRYVAQGKGILQPHHLIDELNNAVCDDTACEKLKEG-PFCEILKSTQEAIVLPPFVA
29289943 -----------MSNPKFTRVPSMRERVEDTLSAHRNQLVALLSRYVAQGKGILQPHHLIDEFNSAVCDDTACEKLKEG-PFCEILKSTQEAIVLPPFVA
115310620 -----------MATIKLQKLPSIRERVEDTLSAHRNELVALLSRYVAQGKGMLQPHHLIDELDNIVVDETACKKLSEG-PFSEVLRSAQEAIVLPPFVA
4468153 -----------MATPKLTKIPSMRERVEVTLAAHRNELVSLLSRYVAQGKGLLQSHHLIDELENIILDDDAKKKLSDG-PFSEVLRSAQEAIVLPPFVA
345286419 -----------MSNPKLTKITSMREGVEDTLSAHRNQLVSLLSRYVAQGKGILQPHHLIDELDNIIDDDSCRAKLNDG-PFGEVLKTAQEAIVLPPFVA
4468151 -----------MTTHEKAPASPCMRERVEETLADHRDELISLLSRYTSRGKCLLQSHQLMDELEDAIAEDPANKKLSDG-CFNEVLSSAQEAIVLPPFVI
151176306 -------------MAPKLTRIPSMRERVEDTLSVHRNELVSLLSKYVAQGKCLLQPHHLIDGLESVIGEDKGKQILSDG-PFSEVLRSAQEAIVVPPFVA
14530225 -------------MAPKLTRIPSMRERVEDTLSVHRNELVSLLSKYVAQGKCLLQPHHLIDGLESVIGEDKGKQILSDG-PFSEVLRSAQEAIVVPPFVA
15239816 ---------------MPTGRFETMREWVYDAISAQRNELLSLFSRYVAQGKGILQSHQLIDEFLKTVKVDGTLEDLNKS-PFMKVLQSAEEAIVLPPFVA
297795665 ---------------MPTGRFETMREWVHDAISAQRNELLSLFSRYVAQGKGILQSHQLIDEFLKTVKVDGTTEDLNKS-PFMKVLQSAEEAIVLPPFVA
225431790 ---------------MPHRYDGQSMRERFQETLSAHRNELVSLFTGYVAQGKGILQPHHMIDELDKVVGKDEGMQKLRDS-PFSKVLKSAQEAIVLPPFVA
255551835 ------------------MPLSSFRDRVHDILSVYRVELVSLLTRHVAEGKGILQTHDLLCELDNVVVDDEAMEKLRRS-PFVEVLQSTQEAIVLPPFVA
168009716 -----------MATQPALRRLNSIQERVQKVVQSNRNLILDLLSRYVKQGRTILQPHHLLDELN-NLGDADQVAEIKDS-AFGNLLQNCQEAMVLPPWVG

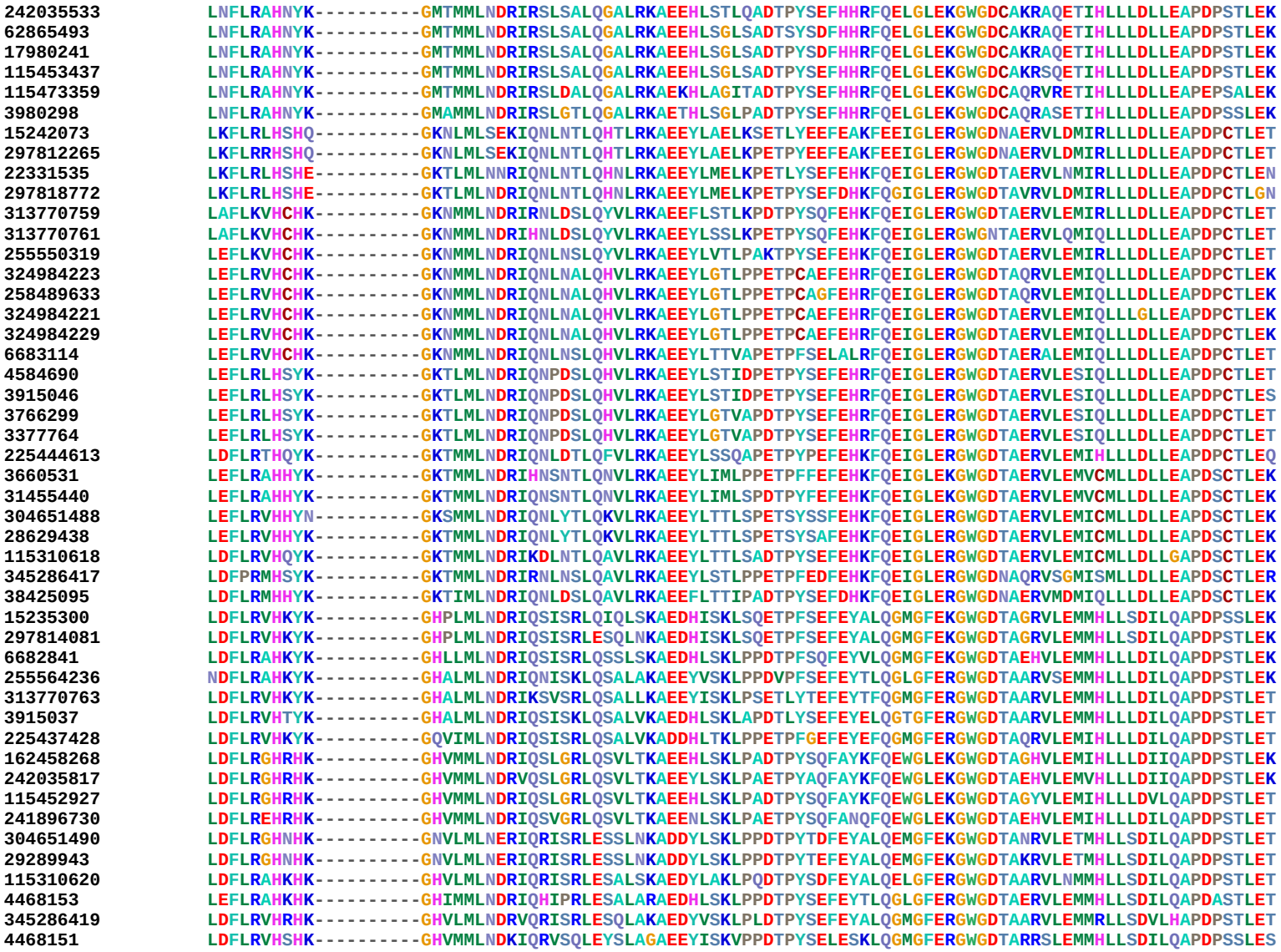
```
168035060          -----MAANGVAPKKPVLQRLNSIQERVQSAVQEHRNVIIDLLSRYVKQGRTHLQPHHIVDELN-SLTEADRVTEIKDS-AFGLLLLNCQEAIVLPPWLG
168058907          ----------MSQPRPTLRRLTSLKERVESSLQEHRNELLHLLQGYVAQGRSILQPHHLQDQLA-AVHDA---AHIQDT-AIGKLLQNCQEAMVSPPWVG
168029793          -------MDGIATQAGALPRMTSMNKKIQGSLDDHRNENLRILSKLTAKRKALMQPHEVIDELN-KAAEESGSLKIMDG-PLARVFSLCQEAIVLAPWVG
94266940           ----------------------MYDDLGSFIDEAEIPLLREFLLSLPPSEPPLLLRNDVLQAFQQFCEAHSEVDEPLG---CWRFLRKVQEILQADDLLV
94264333           ----------------------MYDDLGSFIDEAEIPLLREFLLSLPPSEPPLLLRNDVLQAFQQFCESHPEVDDPLG---CWRFLRKVQEILQADDLLV
297569306          ----------------------MENDLGCFVDEQDVAPLREFLLSLQRPEPPLLLRNDILLAFQQFAGQRPAAEGDLAG--AWRFLRKVQEVLLADDLLI
254416162          ----------------------MLNLIQNVLDCEERNDLRQFASQLKQSGPRYLLRNDILSDFSKYCADQEKPDYFYHSSNLGRLIYYTQEIILDQESLY
332712456          ----------------------MLNLIQDVLESDEKSDLRHFTSQLKTAEPRYLLRNEILAAFNEYCTKHKKSEYFYHSSHLGKLIYYTQEIILEDESLC
17232477           ----------------------MSELMQAILDSEEKHDLRGFISELRQQDKNYLLRNDILNVYAEYCSKCQKPETSYKFSNLSKLIYYTQEIIQEDSNFC
75908500           ----------------------MSELMQAILDSEEKNDLRAFISELRLQDKNYLLRNDILNVYSEYCSKCQKSETSYKFSNLGKLIYYTQEIIQEDSNFC
186682280          ----------------------MSELLQAVLDSEERSDLRSFLSELRQQEKKYLLRNDILNVYSEYCSKSQKPEAYYTTSELGKLIYYTQEIIQEDSNFC
119512682          MQGGGVSFVLPVKNANLMLGVPMSELIQAILDSEEKSDLRSFASQLRQEEKNYLLRNDILNVFIDYCSKSEKSETSAASSRLGKLIYYTQEIIQEDSSLC
334117431          ----------------------MSELFQAVIASDEKTELRQFVSDLRALGNKYLLRNDIVNAFAAYCTKYEKPEQFHQSSLLSKLIYYVQEIILEDGSLC
218440696          ----------------------MSGLIETVLNSEEKIVLWQFIGELRTSDKRYLLRNDIVCAFEDFCKRNGNSQ-KIDSSPLGQMIHYTQEIILEEESIC
159031025          ----------------------MAELIEAIANSNEKKDLQQFLHQLRQLEKPYLLRNDLLTAFADYCQRLDKPASFRQSSLLAKLIYYTQETILESESFC
220907171          ----------------------MSELITAVLASEERRELKEFLLQLQRQDNRYLLRNEILAAFSQYGEEAGKPKEFFHSSQLGKLINFTQEIILEDESFC
22298591           ----------------------MTCVLLKAVVESDERADLRQFSRILQLGEKRYLLRNDILDAFADYCRDQERPVPPPSESRLSKLVFYTQEIIVDNESLC
158339122          ----------------------MSQLIQAVLNSDEKTDLRQFASEVHNQPERYLLRNDILSVFDTFCEKYEKTPAFQLSSRLQKLIYYTQEILVEDENLY
284053161          ----------------------MSELIQNVIESDEKQDLRQFYSELRISEKRYLLRNDILQAFADYCLKFQKAAHFVDSSHLSKLIYYTQEIILENESIC
209523126          ----------------------MSELIQNVIDSDEKQDLRQFASELRISEKRYLLRNDILQAFADYCLKFQKSDHFVHSSNLSKLIYYTQEIILENESIC
17228554           ----------------------MHELFHPIFANGEEKAALQQLIIALDASGKRYFLRNEILHTFSQYCQQAQKPTYFYYSSSVGKLIQYTHEIVLAEDGTW
75909957           ----------------------MHELFHPIFANGEEKAALQQLIIALDSSGKRYFLRNEILHTFSQYCQQAQKPTYFYYSSSVGKLIQYTHEIVLAEDSTW
186685043          ----------------------MYELVQAVFNGDEKTALHQLIYTLSASGKRYLLRNEILQAFADYCHESQKPAYFYHSSSIGKLIQYTHEIIIEEESTW
119510624          ----------------------MHELFETVLNKDEKITLRQFISELSDTDKRYFLRNEILHNFAEFCHQYQKPTYFYYSSSIGRLIHNTHEMILDEQGTW
298489784          ----------------------MYELFQAVLTSDEKTALRQFLLILHDFDKCYFLRNEILQAFTNYCHQSEKPTYFYHSSSLGTLIHYTHEIIIEDENTW
220909283          ----------------------MRDLVESVLTSDEKIDLARFLDRLIQQGKQYFLRNEILQHFSSYCREAEKAAHFYYASYLGKLLHYTHELILEEGAVW
332708740          ----------------------MSSLLRAVLQSNEKADLRQFISQLRSEQERYFLRNQILQAFDNYCTTHDKPAYFKRTSSIAELLNYTHEIILEEKNLW
Clustal Consensus

                                  110       120       130       140       150       160       170       180       190       200
                                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
114331077          FVLRPGIAELEFVRMHQEHLMPEKITINEFLGFKE-TVTKG-EA--IESILEVDFGPFNRAFPKLRESRSIGQGVIFLNRQLSSEMFT-RIE--AGSTRL
30249199           LALRPEIAKWEFMRIHQEHLIPEKLTISEFLKFKE-TVVKG-EA--TESVLEVDFGPFNRGFPRLKESRSIGQGVIFLNRKLSSEMFS-RIE--AGHTSL
82703384           MAAREEAGCWWYARLHLDRLIPEAVTVSEYLAFKE-LLVNPEGA--NEPVLEIDFAPFNRGSPKLKEIRSIGQGVIFLNKQLAGGLFG-QLG--LGSDKL
198283392          LAWRPSPGRWTYLRMHWEQLHLETLAPSDYLAFKE-RQVLPAND--QEPILTVDFEDFRAAPYHLQDEDTIGQGLIYMNRRLAGRLFG-NIK--TGRQSI
344199811          LAWRPSPGRWTYLRLHREQLNLETLTAGDYLAFKE-RQVLPAND--QEPVLTVDFEDFRAVSYRLHDEATIGQGLMYMNRRLAGQLFG-DIK--AGRQSI
255021595          FAWRPRPGRWIYVRIHREQLALEELSTDAYLQAKE-GIVGLGAE--GEAVLTVDFRDFRPVSRRLRDESTIGDGLTHLNRRLAGRIFS-DLA--AGRSQI
77166514           LSVRPRIANWEYYRIHTEVMQIETVPVSQFLEFKERLVLGPTQP--QSWPLKIDMGPFNREFPRLRETRSIGRGMDFLNRHLSNQLFN-ELE--TGGQYL
300115586          LSVRPRIANWEYYRIHTEVMHIETVTVSQFLEFKERLVLGTSQP--QSWPLKIDMGPFNREFPRLKETRSIGRGMDFLNRHLSNQLFN-ELE--TGGQYL
292493898          LSIRPGIASWEYYRIHAEVIQIETVTISQFLEFKARLVLGPQQD--EPWPLKVDMGPFNREFPRLSETRSIGRGMDFLNRHLSSQLFK-ELE--TGGQCL
350552412          FAVRPRVGRWWYLRLHVDDLYPDSVEASEFLGFKERLISAQAG---TERPLEFDIEPFDRGFPKLRESRSIGRGMEFLNRKLSSQLFD-GGD--QGLEKL
220933364          LALRPRIGRWQFLRIHADDLSVEDLGVSEFLAIKERLVCPAPR---HGRPLEFDIEPFNREFPRLRETRSIGRGVEFLNRKLSSQLFD-RAN--GGLDKL
350574485          LALRRRVARWEFVRLHIETMDAQPVSVAEYLAFKERTATGGPE---DPWGLEIDMSPFYRDQFKLREEGSIGRGVEFLNRRLSSRLFE-ELG--KGDRRL
95929190           LAARSTIGHWNYYRLHSEEIEIDEIDVSEYLEFKERLVGYEAPS--DEYLLELDMTPFNREFPKLQEARSIGRGVEFLNRHLSSKLFV-EKR--EGSRKI
298528445          LAVRPEIAKWQYFRFHVEDVLFDEIGASDYLKFDEMQVNNSTQV--DEFLLEIDLEPFNREFPKLNEYTYIGKGVDFLNRHLSGQFFQ-DKK--RGHEKL
297602308          FAVRMNPGIWEYVKVHSDDLSVEGITPSEYLKFKETLYDEKWAKDDNS--LEVDFGALDLSTPHLTLPSSIGNGLQFVSKFMSSKLG--GKP--ESMKPL
115457664          FAVRMNPGIWEYVKVHSDDLSVEGITPSEYLKFKETLYDEKWAKDDNS--LEVDFGALDLSTPHLTLPSSIGNGLQFVSKFMSSKLG--GKP--ESMKPL
```

```
326504012          FAVRMNPGIWEYVKVHADDLSVEGITPSEYLKFKDTLYDEKWAKDDNS--LEVDFGALDLSTPRLTLPSSIGNGMQFVSKFMSSKLN--GKP--ESMKPL
115450038          FAVRMNPGIWEFVKVHSANLSVEQMTPSDYLKNKEALVDDKWGAYDDDSQLEVDFGALDLSTPHLTLPSSIGKGAHLVSRFMSSKLT--DNK-----KPL
326531526          FAVRTNPGVWEFIRVHSGDLSVEQITPADYLKCKETLYDEKWARDDNS--LEVDFGALDLSTPHLALPSSIGNGMQFISRFMSSKLS--GKP--ESMKPL
242063616          LAVRRNPGVWEYITVHSGDLTVQQITPSDYLKRK-----EILFLYDNSSQLEVNLGALDLSTPRLTLPCSIGNGMHLVSRFLSSRLGGGGGR--TKNKAL
42568160           FAVRGTPGNWQYVKVNSSNLSVEALSSTQYLKLKEFLFDENWANDENA--LEVDFGALDFTLPWLSLSSSIGNGLSFVSSKLGGRLN-------DNPQSL
297805240          FAVRGNPGIWQYTKVNSSNLSVEALSSTQYFKLKELLFDENWANDENA--LEVDFGALDFTLPWLSLPSSIGNGVSFVSSKLGSRLN-------DNPQSL
313770765          FSIRPSPGFWEYVKVNSANLSVEGITVTDYLKFKEMIYDENWAKDANA--LEVDFGAFDFSVPHLTLSSSIGNGLGFVSKFVTSKLSG--RL--ENAQPL
313770767          LSIRPSPGFWEFVKVNSADLSVEGITATDYLKFKEMIYDENWAKDANA--LEVDFGAFDFSVPHLTLSSSIGNGLGFVSKFATSKLSG--RL--ESAQPL
255584097          FAIRPSPGFWEFVRVNSADLAVDGINVSEYLKFKEMIFEESWAKDVNT--LEVDFGAFDFSMPKLTLSSSIGNGHNFVSKFITSKLNG--RP--ENAQPL
225466221          FSIRSNPGFWEYVKVSSDDLSVEAITAADYLKFKEMVFDENWAKDDNA--LELNFSAFDFPMPRLTLSSSIGNGVSLVSKFMTSKLNG--NS--QSAQPL
313770769          FAVRPNPGFWEYVKVNAEDLSVEGISVSEYLQLKEMVFDEKWANNENA--LELDFGAMDFSTPRLTLSSSIGNGVNYMSKFMSSKLSG--SS--EAAKPL
313770771          FAVRPNPGFWEYVKVNAEDLSVDGISVSEYLQFKEMIFDEKWASNENA--LEVDFGAMDFSTPRLTLSSSIGNGLNYMSKFMSSKLRG--NS--DAAKPL
255570671          FAVRPNPGFWEYVKVNADDLNVDGISPSEYLQFKEMVFDEKWAKDENA--LEIDFGAIDFSIPRLNLSSSIGNGMSFISKFMSSNLCG--SY--SSAKPL
15219457           LAARPNPGFWEYVKVNSGDLTVDEITATDYLKLKESVFDESWSKDENA--LEIDFGAIDFTSPRLSLSSSIGKGADYISKFISSKLGG--KS--DKLEPL
297842089          LAARPNPGFWEYVKVNSGDLTVDEITANDYLKLKESVFDESWSKDENA--LEIDFGAIDFTSPRLSLSSSIGKGADYISKFISSKLGG--KS--GRLEPL
115466896          LAIRPRPGVWDYIRVNVSELAVEELSVSEYLAFKEQLVDGHTN---SNFVLELDFEPFNASFPRPSMSKSIGNGVQFLNRHLSSKLFQ--DK--ESLYPL
17980243           LAIRPRPGVWDYIRVNVSELAVEELSVSEYLEFKEQLVDGHTN---SNFVLELDFEPFNASFPRPSMSKSIGNGVQFLNRHLSSKLFQ--DK--ESLYPL
162460741          LAIRPRPGVWDYIRVNVSELAVEELSVSEYLAFKEQLVDGQSN---SNFVLELDFEPFNASFPRPSMSKSIGNGVQFLNRHLSSKLFQ--DK--ESLYPL
239984690          LAIRPRTGVWDYIRVNVSELAVEELTVSEYLAFKEQLVDEHAS---SKFVLELDFEPFNASFPRPSMSKSIGNGVQFLNRHLSSKLFQ--DK--ESLYPL
162460681          LAIRPRPGVWEYVRVNVSELAVEELRVPEYLQFKEQLVEEGPN---NNFVLELDFEPFNASFPRPSLSKSIGNGVQFLNRHLSSKLFH--DK--ESMYPL
242035533          LAIRPRPGVWEYVRVNVSELAVEELRVPEYLQFKEQLVEEGPN---NNFVLELDFEPFNASFPRPSLSKSIGNGVQFLNRHLSSKLFH--DK--ESMYPL
62865493           LAIRPRPGVWEYVRVNVSELAVEELRVPEYLQFKEQLVEGSTN---NNFVLELDLEPFNASFPRPSLSKSIGNGVQFLNRHLSSKLFH--DK--ESMYPL
17980241           LAIRPRPGVWEYVRVNVSELAVEELRVPEYLQFKEQLVEGSTN---NNFVLELDFEPFNASFPRPSLSKSIGNGVQFLNRHLSSKLFH--DK--ESMYPL
115453437          LAIRPRPGVWEYVRVNVSELAVELLTVPEYLQFKEQLVEEGTN---NNFVLELDFEPFNASFPRPSLSKSIGNGVQFLNRHLSSKLFH--DK--ESMYPL
115473359          LAIRPRPGVWEYLRINVSQLGVEELSVPEYLQFKEQLVDGSTQ---NNFVLELDFEPFNASFPRPSLSKSIGNGVQFLNRHLSSKLFH--DK--ESMYPL
3980298            LAIRPRPGVWEYVRVNVSELGVEELSVLRYLQFKEQLANGRTD---NNFVLELDFGPFNASFPRPSLSKSIGNGVQFLNRHLSSKLFH--DK--ESMYPL
15242073           LAVRPRPGVWEYLRVNLHALVVEELQPAEFLHFKEELVDGVKN---GNFTLELDFEPFNASIPRPTLHKYIGNGVDFLNRHLSAKLFH--DK--ESLLPL
297812265          LAVRPRPGVWEYIRVNLHALVVDELQPAEFLHFKEELVDGVKN---GDFTLELDFEPFNASFPRPTLNKYIGNGVEFLNRHLSAKLFH--DK--ESLLPL
22331535           LAVRPRPGVWEYVRVNLHDLVVEELQASEYLQFKEELVDGIKN---GNFTLELDFEPFNAAFPRPTLNKYIGDGVEFLNRHLSAKLFH--DK--ESLHPL
297818772          LAVRPRPGVWEYVRVNLHDLVVDELQASEYLQFKEELVDGIRN---GNFTLELDFEPFNAAFPRPTLNKYIGNGVEFLNRHLSAKLFH--DK--ESLHPL
313770759          LALRPRPGVWEYIRLNVQALVVEELRVAEYLHFKEELVDGGCN---GNFVLELDFEPFNASFPRPTLSKYIGNGVEFLNRHLSAKLFH--DK--ESLHPL
313770761          LAVRPRPGVWEYVRVNVQALVVEELRVAEYLHFKEELVDGGSN---GNFVLELDFEPFSASFPRPTLSKYIGNGVEFLNRHLSAKLFH--DK--ESLHPL
255550319          LAVRPRPGVWEYIRVNVHALVVEELRVAEYLHFKEELVDGSQN---GNFVLELDFEPFNASFPRPTLSKYIGNGVEFLNRHLSAKLFH--DK--ESLHPL
324984223          LAVRPRPGVWEYIRVNVHALVVEELTVAEYLHFKEELVDGSSN---GNFVLELDFEPFNSSFPRPTLSKSIGNGVEFLNRHLSAKLFH--DK--ESMHPL
258489633          LAVRPRPGVWEYIRVNVHALVVEELTVAEYLHFKEELVDGSSN---GNFVLELDFEPFNSSFPRPTLSKSVGNGVEFLNRHLSAKLFH--DK--ESMHPL
324984221          LAVRPRPGVWEYIRVNVHALVVEELTVAEYLHFKEELVDGSSN---GNFVLELDFEPFNSSFPRPTLSKSIGNGVEFLNRHLSAKLFH--DK--ESMHPL
324984229          LAVRPRPGVWEYIRVNVHALVVEELTVAEYLHFKEELVDGSSN---GNFVLELDFEPFNSSFPRPTLSKSIGNGVEFLNRHLSAKLFH--DK--ESMHPL
6683114            LAVRPRPGVWEYIRVNVHALVVEELLVAEYLHFKEELVDGGSN---GNFVLELDFEPFNASFPRPTLSKSIGNGVEFLNRHLSAKLFH--DK--ESMHPL
4584690            LAVRPRPGIWEYLRVNVHALVVENLQPAEFLKFKEELVDGSAN---GNFVLELDFEPFTASFPRPTLNKSIGNGVQFLNRHLSAKLFH--DK--ESLHPL
3915046            LAVRPRPGIWEYLRVNVHALVVENLQPAEFLKFKEELVDGSAN---GNFVLELDFEPFTASFPRPTLNKSIGNGVHFLNRHLSAKLFH--DK--ESLHPL
3766299            LAVRPRPGVWEYLRVNVHALVVENLQPAEFLKFKEELVDGSAN---GNFVLELDFEPFTASFPRPTLNKSIGNGVQFLNRHLSAKLFH--DK--ESLHPL
3377764            LAVRPRPGVWEYLRVNVHALVVENLQPAEFLKFKEELVDGSAN---GNFVLELDFEPFTASFPRPTLNKSIGNGVQFLNRHLSAKLFH--DK--ESLHPL
225444613          FAVRPRPGVWEYIRVNVSALVVEELLVPEYLHFKEELVDGSCN---GNFVLELDFEPFTASVPRPTLSKSIGNGVEFLNRHLSAKMFH--DK--DSMQPL
3660531            LAIRLRPGVWEYVRVNVNALVVEELSVPEYLQFKEELVDGASN---GNFVLELDFEPFTASFPKPTLTKSIGNGVEFLNRHLSAKMFH--DK--ESMAPL
31455440           LAIRLRPGVWEYIRVNVNALVVEELSVPEYLQFKEELVDGASN---GNFVLELDFEPFTASFPKPTLTKSIGNGVEFLNRHLSAKMFH--DK--ESMTPL
304651488          LAIRLRPGVWEYVRVNVNALSVEELTVPEFLQFKEELVNGTSS---DNFVLELDFEPFTASFPKPTLTKSIGNGVEFLNRHLSAKMFH--DK--ESMTPL
28629438           LAIRLRPGVWEYVRVNVNALIVEELTVPEFLQFKEELVNGTSN---DNFVLELDFEPFTASFPKPTLTKSIGNGVEFLNRHLSAKMFH--DK--ESMTPL
```

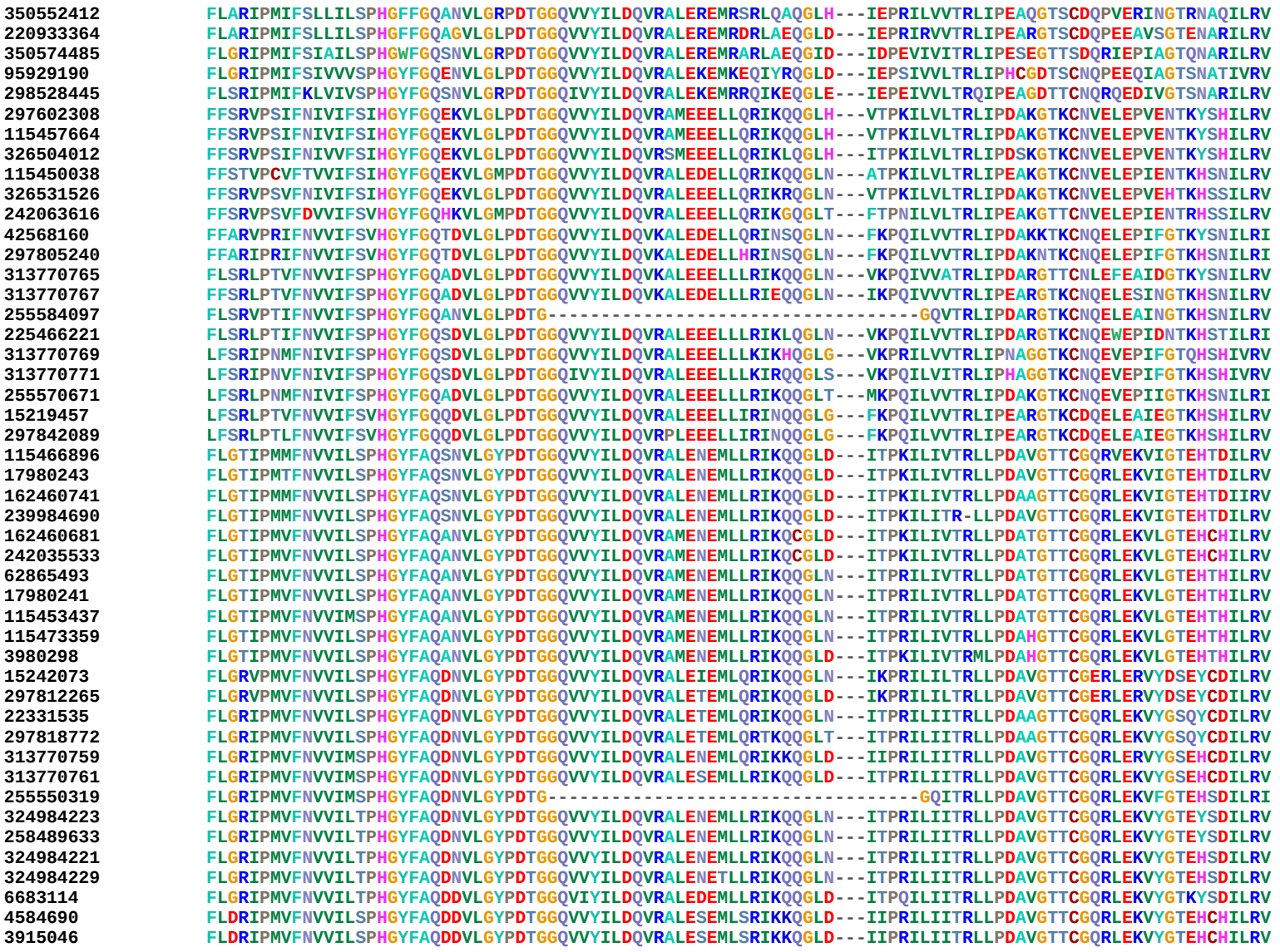
```
115310618          LAIRLRPGVWEYVRVNVHALVVEELTVPEYLHFKEELVDGSKN---GNFVLELDFEPFTASFPKPTLTKYIGDGVEFLNRHLSAKMFH--DK--ESMAPL
345286417          LAIRLRPGVWEYVRVNVNALVVEELTVPQYLHFKEELVNGAAN---GNFVLELDFEPFTASFPKPTLTKSIGNGVEFLNRHLSAKMFH--DR--ESMTPL
38425095           LAVRPRPGIWEYIRVNVDALAVEELTPSQFLHVKEELVDGSTN---GNFVLELDFEPFTASFPRPTLSKSIGNGVEFLNRHLSAKMFH--DK--ESMRPL
15235300           LAVRPRPGVWEYVRVNVFELSVEQLTVSEYLRFKEELVDGPNS---DPFCLELDFEPFNANVPRPSRSSSIGNGVQFLNRHLSSVMFR--NK--DCLEPL
297814081          LAVRPRPGVWEYVRVNVFELSVEQLTVSEYLRFKEELVDGPNS---DPFPLELDFEPFNANVPRPSRSSSIGNGVQFLNRHLSSVMFR--NK--DCLEPL
6682841            IAVRPRPGVWEYVRVNVYELSVEQLSVSEYLHFKEELVDAAFN---ERFVLELDFEPFNATFPRPNRSSSIGNGVQFLNRHLSSSMFR--NK--DCLEPL
255564236          IAIRPRPGIWEYVRVNVHDLSVEQLDVSQYLRFKEELVDGSSN---DPYVLELDFEPFNADVPKPHRSSSIGNGVQFLNRHLSSIMFR--NK--DCLEPL
313770763          VSIRPRPGVWEYVRVDVSQLNVEELTVSQYLRFKEELVDGPSN---DPYVLELDFEPFNAAFPRPTRSSSIGNGVQYLNRHLSSNMFR--NK--DTLEPL
3915037            IAVRPRPGVWEYVRVNVFELSVEQLSVSEYLSFKEELVEGKSN---DNIILELDLEPFNASFPRPTRSSSIGNGVQFLNRHLSSNMFR--NK--DCLEPL
225437428          IAVRPRPGVWEYVRVNVHELSVDQLSVSEYLRFKEELVDGMFN---DYYVLELDFEPFNASFPRPNRSSSIGNGVQFLNRHLSSIMFR--NK--ESLEPL
162458268          IAVRPRPGVWEYVRVNVHELSVEQLTVSEYLRFKEELVDGQHN---DPYVLELDFEPFNVSVPRPNRSSSIGNGVQFLNRHLSSIMFR--NR--DCLEPL
242035817          IAVRPRPGVWEYVRVNVHELSVEQLTVSEYLRFKEDLVDGQHN---DPYILELDFEPFNASVPRPNRSSSIGNGVQFLNRHLSSIMFR--NR--DCLEPL
115452927          IAVRPRPGVWEYVRVNVHELSVEQLTVSEYLRFKEELVDGQYN---DPYILELDFEPFNASVPRPNRSSSIGNGVQFLNRHLSSIMFR--NK--DCLEPL
241896730          IAVRPRPGVWEYVRVNVHELNVEQLSVSEYLRFKEELVDGQHN---NPYVLELDFEPFTALIPRPSRSSSIGNGVQFLNRHLSSILFR--NR--DCLEPL
304651490          IAVRPRPGVWEYVRVNVYDLSVEQLTVPEYLRFKEELVDGEDHN--HLFVLELDFEPFNASVPRPSRSSSIGNGVQFLNRHLSSNMFR--SN--ESLDPL
29289943           IAVRPRPGVWEYVRVNVYDLSVEQLTIPEYLRFKEELVDGEDN---NLFVLELDFEPFNASVPRPSRSSSIGNGVQFLNRHLSSNMFR--SK--ESLDPL
115310620          IAVRPRPGVWEYVRVNVYELSVDQLSISEYLHLKEELVDGRSE---DHLVLELDFEPFNATFPRPTRSSYIGNGVQFLNRHLSSIMFR--NK--DSLEPL
4468153            LAVRPRPGVWEFVRVNVYQLSVDELTISEYLRFKEELVDGGID---DNFVLELDFEPFNASFPRPTRSSYIGNGVQFLNRHLSSIMFR--NK--DCLEPL
345286419          IAIRPRPGVWEFVRVNVYELSVDELTVSEYLRFKEALVDGQHD---DHYVLELDFEPFNATFPRPTRSSSIGNGVQFLNRHLSSIMFR--NK--ESLDPL
4468151            LAVRPRPGVWAYVRVNMQELSIDELTVSEYLCFKEELVDGRGF---DPYVLELDFEPFNATFPRPSRSSSIGNGVQFLNRHLSSIMFR--NK--DSLEPL
151176306          IAVRPRPGVWEYVRVNVSELNVEQLTVSEYLHFKEELVDGKAD---DHYVLELDFEPFNESVPRPTRSSSIGNGVQFLNRHLSSSMFR--NK--DCLEPL
14530225           IAVRPRPGVWEYVRVNVSELNVEQLTVSEYLHFKEELVDGKAD---DHYVLELDFEPFNESVPRPTRSSSIGNGVQFLNRHLSSSMFC--NK--DCLEPL
15239816           LAIRPRPGVREYVRVNVYELSVDHLTVSEYLRFKEELVNGHAN---GDYLLELDFEPFNATLPRPTRSSSIGNGVQFLNRHLSSIMFR--NK--ESMEPL
297795665          LAIRPRPGVREYVRVNVYELSVDHLTVSEYLRFKEELVNGHAN---GDYLLELDFEPFNATLPRPTRSSSIGNGVQFLNRHLSSIMFR--NK--ESMEPL
225431790          IAIRPRPGVWEYIRVNVYELNVDQLSVSEYLQFKEELVDGQIK---GNYVLELDFEPFNATFPRPTRSSSIGNGVQFLNRHLSSIMFR--NK--ESLEPL
255551835          MAIRPRPGVWEYVRVNVYELSVDHLNVSEFLRFKEDLADGECD---ESYVLELDFEPFNATFPRPTRSSSIGNGVQFLNRHLSSVMFR--QK--ESLEPL
168009716          FAVRPRPGIWEYVRINVEELTLEELSVSEYLSFKEQLAN-GTEY--DPFVLELDFAPFNANFPHMTRPSSIGHGVQFLNRHLSSKLFH--TP--DSMEPL
168035060          LAVRPRPGIWEYLRINVEELILEELSVSEYLGFKEQLANSTDVR--DPFLLELDMAPFNSNFPRMTRPSSIGHGVEFLNRHLSLKLFQ--TA--DGIEPL
168058907          FAVRPRPGIWEYVRINVEELIVEELSVSEYLGFKEQLSLGSDSI--DLYVLELDFEPFNAHFPRMTRPSSIGHGVQFLNRHLSSKLFQ--NP--ESMEPL
168029793          LALRPRPGLWEYMRINVEEMIVEELTTSEYLSFKECLADENRCN--DLYVLELDIEPFNVGFPRMTRPQSIGNGVQFLNRHLSSRLFR--DA--DSMEPL
94266940           IVYRHRRANCLIFAINGQSDKLLKLSVGDFLAIKERLLRPELPP--QQRTLNLNLAPFYDYGPTLKDPNTIGQGIKFLNRHMSGRLANHPEK---WNRFL
94264333           IVYRHRRANCLIFAINGQSDKLLKLSVGDFLAIKERLLRPELPP--QQRTLNLNLAPFYDYGPTLKDPNTIGQGIKFLNRHMSGRLANHPEK---WNRFL
297569306          ILYRHRRASCRIFALSEDDDRLIPVTLIDFLAIKERLIRPNLPP--QQRTLAINLAPFYDYGPTLKDPNTIGQGIKFLNRHMSGNLSNHPEK---WNRSL
254416162          LIIRPKIAEQEAF-QVLEDLTVEPITIQTLLDVRDRFVNHYRPN--EGDVLELDFKPFYDYSPIIRDSKNIGKGVQFLNRFLSSKLFQDPRQ---WQESL
332712456          LIIRPKIAAKRAF-RLFEDLRAQEVTPEELLDIGDRFVNRYNPK--VGEVLQLDFQPFYDYSPLIRDPKNIGKGVRFLNRYLSSKLFQDPEQ---WLESL
17232477           FIIRPKIAAQEVY-RLTADLDVEPMTVQELLDLRDRLVNKFHPY--EGDILELDFGPFYDYTPTIRDPKNIGKGVQYLNRYLSSKLFQDSQQ---WLESL
75908500           FIIRPNIAAQEVY-RLTADLDVEPMTVQELLDLRDRLVNKFHPH--EGDLLELDFGPFYDYTPTIRDPKNIGKGVQYLNRYLSSKLFQDSQQ---WLESL
186682280          FIIRSKIASQEVY-WLTSDLSIEPMTVQDLLDLRDRLVNKFHPN--DGDLLELDFGPFYDYTPIIRDPKNIGKGVQFLNRYLSSKIFQDSKQ---LLENL
119512682          FIIRPKIASQEVY-RLTEDLNVEPMSVQELLDLRDRFVNKYHPL--EGDLLELDFGPFYDYTPVIRDPKNIGKGVQFLNRYLSSKIFQDSKQ---LLDSL
334117431          VLLRPKIASMEIV-RVADDLTVEQMTVQELLDVRDRFVNHFHPQ--EGDILELDFGPFYDYSPTIRDPKNIGKGVQFLNRYLSSKLFQDARQ---WQETL
218440696          IVYRPKIARQEVY-RLREDTPIEHLSVQQLLDVRDHFVNQFHPD--EGDVFEIDFEPFYDYSPTIRDSKNIGKGVRFLNRYMSSKLFQDPRQ---WLETL
159031025          LVVRPQIALSQAF-RLTDDLTCEPISVQELLDLRDRLVHRSHPE--EGDLLELDFRPFYDYSPIIRDAKNIGKGVQFLNRFLSSRLFQAEQTGEQWLQSA
220907171          LVLRPNIASQSVF-RITSDLAVEEMRVKELLTVRDRLVQRHHPE--EGELLELDFQPFYDYSPTIRDAKNIGKGMQLLTRYLSSKLFQDPQE---WTQVL
22298591           WIVRPRIAQQEVCRLLVEDLTIVPMTIPELLDLRDRLVNHYHPN--EGDVFEIDVQPFYDYSPIIRDAKNIGKGVEFLNRYLSSKLFQDPRQ---WQQNL
158339122          LIIRPKIASEEAYRLDPRELVYEQIQIDELLDLRDRFVGHYHPQ--EGDILEIDFRPFYDYSPVIRDPKNIGRGVQYLNRYLSSKMFEGPQQ---WLFSL
284053161          LIVRPQIATQERF-RLRPDLSVDVMSVQELLDLRDHFVDHHHPH--DGNVFAIDFGPFYDYSPQLRDSKNIGKGVRYLNRYLSSKLFQDSNQ---WLESL
209523126          LIVRPQIATQERF-RLRPDLSVDVMSVQELLDLRDHFVDHHHPH--EGNIFAIDFGPFYDYSPQLRDSKNIGKGVRYLNRYLSSKLFQDSNQ---WLESL
```


17228554 FVVRPRIASQEVWRLTSDLAKFDSMSIDAFLDVSDRLVNAYEP-----NILEIDLNSFYEASPSISDPRNIGQGLAFLNRYLCSQIATDPQY---WVELV
75909957 FVVRPRIANQEVWRLTSDLAKFDSMPIDAFLDVSDRLVNAYEP-----NILEIDLSSFYEASPSISDPRNIGQGLAFLNRYLCSQIATDPQY---WVELV
186685043 FVIRPKIANQEVWRLTANLDSFEQMTQQALLDVRDRLVNRYQP-----GILEIDLHPFYEDSPRIDDPRNIGQGLAYLNRYLCNQLLTDPEY---WVEML
119510624 FVVRPRIGSQQMWRLQADFSGFEPMTPQAWLDVSDRLVNRYQP-----HILEIDFQPFAEESTRITDPRNIGQGLAFLNRYLCDQLSNDTHY---WLEVI
298489784 FVVRPRIASQEVWRLTADFTHFDLMTPKAFLDVSDRLVNRYQP-----HILEIDLHPFYQTSPRISDPREIGQGLTFLNHYLCNQFVSDPQY---WLETF
220909283 LLVRSTINDQQIWQISTDLNRYGRMSPHDLLEVRDRLVNRSQS-----SILEINVHPFYNMAYAVEDPRNIGQGLAFLNHYLCNQVSIDRTR---WLDVL
332708740 LLLRTRVASQEIYRLAADLTSFEPMPVEELLSLRDRWVKRYFPE--KGGLLEIDVGPFYKNTPTIRDPRKIGNGLEFLNRYLSSQLFADSEQ---WLEEL
Clustal Consensus * : : : .:. : :* * :. .
210 220 230 240 250 260 270 280 290 300
114331077 LHFLGVHTID----------GQQLMFTSNSHNINMVRSQLRQALEMLEAVDGTTPWAELSSDMSKIGFAPGWGHNAARVAETMNMLMDILEAPSPSALEA
30249199 LHFLGVHAIE----------GQQLMFSNNSHDIHAVRNQLRQALEMLETLDGTTPWIELAPKMNQLGFAPGWGHNANRVAETMNMLMDILEAPSPSALEE
82703384 LHFLTVHSMD----------GKQLMLGGNFADVPALRSGLRRALSMLEKYPDDTEWKDVAEPLGGIGFAPGWGNCVGRVSETMSLLVDILEAPSPQILES
198283392 LDFLAVHKLN----------GQSLMVHDQPPDFEALR----QTVQYLATLPKTKPWTEFAAEMTYRGFAPGWGDTAGRVRETMRLLMDLLDAPSAEGLQA
344199811 LDFLAVHKLN----------GQSLMVHDQPPDFEALR----RTVQYLATLPKTQAWTEIAAEMTHRGFAPGWGDTVGRVRETMRLLMDLLDAPSAEGLQA
255021595 LEFLSLHRLD----------GQNLMLSNGNTDFDSLR----QTVQYLGTLPRETPWAEIREDMRRRGFAPGWGNTAGRVRETMRLLMDLLDSPSPAALES
77166514 LSFLSVHHCR----------GQPLMLNDRIQDVQGLRCALRLAMDFLGGFQEAAEWDAVGHKLQEFGFERGWGRTAARIQDSFSLLMDILEAPEPGNLEH
300115586 LSFLSVHHCR----------GQPLMLNDRIQDVRGLRRALRLAMDFLGSFQEAAEWDAVGHKLQELGFERGWGRTAVRMQDSFSLLMDILEAPEPGNLEH
292493898 LNFLSVHHCR----------GQPLMLNDRIQDLRGLRRALRRAVDFLGGFPKAAEWEAVGHKLQELGFERGWGGTVAQMEDSFSLLMDILEAPDPGNLER
350552412 FHFLHEHCCE----------GQILMINERVRNVNELRGVIRRCEKLLNGYARHTPWAEVAPLLRDVGLEPGWGSDVGRILETLRLLSDLLEAPSPETLER
220933364 FRFLREHRCD----------GRLLMINDRIRDVDALRVAIRDAEQRLARLKRDTPWADFAHPLQDLGFEPGWGRDAGRVQETLRLLSELLEAPSPESLER
350574485 LNFLRMHSHR----------GQVLMLNDTITDVAGLRNALRQALLPLRRRAASTPYEELAPDLRPLGFEPGWGCDAARVRNTMGLLLDILEAPSPQTIEE
95929190 LDFLRVHQHR----------STQLMLNGMIEDVPGLQAALRKGVKFLKNCDEDTCWDDVAPTMMSYGFQPGWGRTLEDILEMFHMLMDILEAPDPQNLEK
298528445 FEFLRLHQIE----------GKQLMLNGHIETVSGLRSALRKALTFLKKQDPSQKWSGISRHMQTLGFQPGWGKDVERVRENLELLREILEAPTPNILAS
297602308 LDYLLTLNYR----------GEKLMINDTIDTVSKLQTALLLAEVFVSGLPKYTPYLKFEQRFQEWGLEKGWGDTAERCKETLNCLSEVLQAPDPTNMEK
115457664 LDYLLTLNYR----------GEKLMINDTIDTVSKLQTALLLAEVFVSGLPKYTPYLKFEQRFQEWGLERGWGDTAERCKETLNCLSEVLQAPDPTNMEK
326504012 LDYLLALNYR----------GEKLMVNDTIDTVNKLQTALLLAEVFVSGLPKFTPYLKFEQRFQEWGLEKGWGENAERCKETLNFLSEVLQAPDPINMEK
115450038 LDYLLALSHR----------GDKLMINDILDTVDKLQTALLLAEVYVAGLHPDTNYSEFEQKFQEWGLEKGWGDTAETCKETLSSLSEVLQAPDPINMEK
326531526 LDYLLALNYR----------GEKLMISDSLDTADKLQTALLLAEVFVASLEKSTPYQQFEQKFQEWGLEKGWGDTAETCRETLNFLSEVLQAPDPINMEK
242063616 LDYLLALRYYRRRPGDQQQINNKLLISDTLDTVGKLQAALLLAQAFVSEQHPDTPYQQMAHRFQEWGLEKGWGDTAEACGHTLACLAEVLQAPDPASIHR
42568160 VDYLLSLEHQ----------GEKLMMNETLNTARKLEMSLILADVFLSELPKDTPFQAFELRFKECGFEKGWGESAGRVKETMRILSEILQAPDPQNIDR
297805240 VDYLLSLEHQ----------GEKLMMNETLNTARKLEMSLILADVFLSELPKDTPFQAFELRFKEWGFEKGWGESAGRVKETMRILSEILQAPDPHNIDR
313770765 VDYLLSLNRQ----------GEKLMINETLGTVGKLQMALIVAEVYLSGLAKDTPYQNFEISFKEWGFEKGWGDTAERVKETMRCLSEVLQAPDPMNMEK
313770767 VDYLLSLNHE----------GEKLMINETLSSVRKLRMALIVAEAYLSGLPKDTQYQNFETSFKAWGFEKGWGNTAERVKETMRCLSEVLQAPDPLNMEN
255584097 VDYLLSLTHH----------GEKLMINENLSTVAKLQMALIVAEVYLSGLAGDTPYQNFELSFKEWGFEKGWGDTAERAKETMRSLSEVLQAPDPVNMEK
225466221 VDYLLSLNHQ----------GEKLMITNTLNTPTKLQMALIVAEVFVSALPKDTPYPSFELRFKEWGFEKGWGNTAERVKETMRSLSEALEAPDPMNMEK
313770769 LDYLLALNHQ----------GENLMINQTLDTVAKLQEALIVAEVVVSAFPKDTPYQDFQQRLRELGFETGWGDTAERVKETMRLLSESLQAPYPMKLQL
313770771 LDYLLALDHQ----------GENLMINQALDSVSKLQAALIVAEVVVSAFPKDAPYQDFQQSLKRLGFEKGWGDTAERVKETMRMLSESLQAPEPVKLEL
255570671 LDYLLALNYQ----------GEELMINEKLDTVAKLQKALTGAEDVLSVFSKEAAYKNVQQSLKEMGFEKGWGNTAERVKETMRLLSESLQAPDPAKLEL
15219457 LNYLLRLNHH----------GENLMINDDLNTVAKLQKSLMLAVIVVSTYSKHTPYETFAQRLKEMGFEKGWGDTAERVKETMIILSEVLEAPDNGKLDL
297842089 LNYLLRLNHH----------GENLMINDDLNTVAKLQKSLMLAVIVVSTYPKHTPYETFALRLKEMGFEKGWGDTAERVKETMVMLSEVLEAPDNVKLDL
115466896 LNFLKAHNHK----------GTTMMLNDRIQSLRGLQSSLRKAEEYLMGIPQDTPYSEFNHRFQELGLEKGWGDCAKRVLDTIHLLLDLLEAPDPANLEK
17980243 LNFLKAHNHK----------GTTMMLNDRIQSLRGLQSALRKAEEYLMSFPQDTPYSEFNHRFQELGLEKGWGDTAKRVLDTIHLLLDLLEAPDPANLEK
162460741 LNFLKAHNYK----------GTTMMLNDRIQSLRGLQSSLRKAEEYLLSVPQDTPYSEFNHRFQELGLEKGWGDTAKRVLDTLHLLLDLLEAPDPANLEK
239984690 LNFLKAHNYK----------GTTMMLNDRIQSLRGLQSALRKAEEYLISIPEDTPSSEFNHRFQELGLEKGWGDTAKRVHDTIHLLLDLLEAPDPASLEK
162460681 LNFLRAHNYK----------GMTMMLNDRIRSLSALQGALRKAEEHLSTLQADTPYSEFHHRFQELGLEKGWGDCAKRAQETIHLLLDLLEAPDPSTLEK


242035533 LNFLRAHNYK----------GMTMMLNDRIRSLSALQGALRKAEEHLSTLQADTPYSEFHHRFQELGLEKGWGDCAKRAQETIHLLLDLLEAPDPSTLEK
62865493 LNFLRAHNYK----------GMTMMLNDRIRSLSALQGALRKAEEHLSGLSADTSYSDFHHRFQELGLEKGWGDCAKRAQETIHLLLDLLEAPDPSTLEK
17980241 LNFLRAHNYK----------GMTMMLNDRIRSLSALQGALRKAEEHLSGLSADTPYSDFHHRFQELGLEKGWGDCAKRAQETIHLLLDLLEAPDPSTLEK
115453437 LNFLRAHNYK----------GMTMMLNDRIRSLSALQGALRKAEEHLSGLSADTPYSEFHHRFQELGLEKGWGDCAKRSQETIHLLLDLLEAPDPSTLEK
115473359 LNFLRAHNYK----------GMTMMLNDRIRSLDALQGALRKAEKHLAGITADTPYSEFHHRFQELGLEKGWGDCAQRVRETIHLLLDLLEAPEPSALEK
3980298 LNFLRAHNYK----------GMAMMLNDRIRSLGTLQGALRKAETHLSGLPADTPYSEFHHRFQELGLEKGWGDCAQRASETIHLLLDLLEAPDPSSLEK
15242073 LKFLRLHSHQ----------GKNLMLSEKIQNLNTLQHTLRKAEEYLAELKSETLYEEFEAKFEEIGLERGWGDNAERVLDMIRLLLDLLEAPDPCTLET
297812265 LKFLRRHSHQ----------GKNLMLSEKIQNLNTLQHTLRKAEEYLAELKPETPYEEFEAKFEEIGLERGWGDNAERVLDMIRLLLDLLEAPDPCTLET
22331535 LKFLRLHSHE----------GKTLMLNNRIQNLNTLQHNLRKAEEYLMELKPETLYSEFEHKFQEIGLERGWGDTAERVLNMIRLLLDLLEAPDPCTLEN
297818772 LKFLRLHSHE----------GKTLMLNDRIQNLNTLQHNLRKAEEYLMELKPETPYSEFDHKFQGIGLERGWGDTAVRVLDMIRLLLDLLEAPDPCTLGN
313770759 LAFLKVHCHK----------GKNMMLNDRIRNLDSLQYVLRKAEEFLSTLKPDTPYSQFEHKFQEIGLERGWGDTAERVLEMIRLLLDLLEAPDPCTLET
313770761 LAFLKVHCHK----------GKNMMLNDRIHNLDSLQYVLRKAEEYLSSLKPETPYSQFEHKFQEIGLERGWGNTAERVLQMIQLLLDLLEAPDPCTLET
255550319 LEFLKVHCHK----------GKNMMLNDRIQNLNSLQYVLRKAEEYLVTLPAKTPYSEFEHKFQEIGLERGWGDTAERVLEMIRLLLDLLEAPDPCTLET
324984223 LEFLRVHCHK----------GKNMMLNDRIQNLNALQHVLRKAEEYLGTLPPETPCAEFEHRFQEIGLERGWGDTAQRVLEMIQLLLDLLEAPDPCTLEK
258489633 LEFLRVHCHK----------GKNMMLNDRIQNLNALQHVLRKAEEYLGTLPPETPCAGFEHRFQEIGLERGWGDTAQRVLEMIQLLLDLLEAPDPCTLEK
324984221 LEFLRVHCHK----------GKNMMLNDRIQNLNALQHVLRKAEEYLGTLPPETPCAEFEHRFQEIGLERGWGDTAERVLEMIQLLLGLLEAPDPCTLEK
324984229 LEFLRVHCHK----------GKNMMLNDRIQNLNALQHVLRKAEEYLGTLPPETPCAEFEHRFQEIGLERGWGDTAERVLEMIQLLLDLLEAPDPCTLEK
6683114 LEFLRVHCHK----------GKNMMLNDRIQNLNSLQHVLRKAEEYLTTVAPETPFSELALRFQEIGLERGWGDTAERALEMIQLLLDLLEAPDPCTLET
4584690 LEFLRLHSYK----------GKTLMLNDRIQNPDSLQHVLRKAEEYLSTIDPETPYSEFEHRFQEIGLERGWGDTAERVLESIQLLLDLLEAPDPCTLET
3915046 LEFLRLHSYK----------GKTLMLNDRIQNPDSLQHVLRKAEEYLSTIDPETPYSEFEHRFQEIGLERGWGDTAERVLESIQLLLDLLEAPDPCTLES
3766299 LEFLRLHSYK----------GKTLMLNDRIQNPDSLQHVLRKAEEYLGTVAPDTPYSEFEHRFQEIGLERGWGDTAERVLESIQLLLDLLEAPDPCTLET
3377764 LEFLRLHSYK----------GKTLMLNDRIQNPDSLQHVLRKAEEYLGTVAPDTPYSEFEHRFQEIGLERGWGDTAERVLESIQLLLDLLEAPDPCTLET
225444613 LDFLRTHQYK----------GKTMMLNDRIQNLDTLQFVLRKAEEYLSSQAPETPYPEFEHKFQEIGLERGWGDTAERVLEMIHLLLDLLEAPDPCTLEQ
3660531 LEFLRAHHYK----------GKTMMLNDRIHNSNTLQNVLRKAEEYLIMLPPETPFFEFEHKFQEIGLEKGWGDTAERVLEMVCMLLDLLEAPDSCTLEK
31455440 LEFLRAHHYK----------GKTMMLNDRIQNSNTLQNVLRKAEEYLIMLSPDTPYFEFEHKFQEIGLEKGWGDTAERVLEMVCMLLDLLEAPDSCTLEK
304651488 LEFLRVHHYN----------GKSMMLNDRIQNLYTLQKVLRKAEEYLTTLSPETSYSSFEHKFQEIGLERGWGDTAERVLEMICMLLDLLEAPDSCTLEK
28629438 LEFLRVHHYK----------GKTMMLNDRIQNLYTLQKVLRKAEEYLTTLSPETSYSAFEHKFQEIGLERGWGDTAERVLEMICMLLDLLEAPDSCTLEK
115310618 LDFLRVHQYK----------GKTMMLNDRIKDLNTLQAVLRKAEEYLTTLSADTPYSEFEHKFQEIGLERGWGDTAERVLEMICMLLDLLGAPDSCTLEK
345286417 LDFPRMHSYK----------GKTMMLNDRIRNLNSLQAVLRKAEEYLSTLPPETPFEDFEHKFQEIGLERGWGDNAQRVSGMISMLLDLLEAPDSCTLER
38425095 LDFLRMHHYK----------GKTIMLNDRIQNLDSLQAVLRKAEEFLTTIPADTPYSEFDHKFQEIGLERGWGDNAERVMDMIQLLLDLLEAPDSCTLEK
15235300 LDFLRVHKYK----------GHPLMLNDRIQSISRLQIQLSKAEDHISKLSQETPFSEFEYALQGMGFEKGWGDTAGRVLEMMHLLSDILQAPDPSSLEK
297814081 LDFLRVHKYK----------GHPLMLNDRIQSISRLESQLNKAEDHISKLSQETPFSEFEYALQGMGFEKGWGDTAGRVLEMMHLLSDILQAPDPSTLEK
6682841 LDFLRAHKYK----------GHLLMLNDRIQSISRLQSSLSKAEDHLSKLPPDTPFSQFEYVLQGMGFEKGWGDTAEHVLEMMHLLLDILQAPDPSTLEK
255564236 NDFLRAHKYK----------GHALMLNDRIQNISKLQSALAKAEEYVSKLPPDVPFSEFEYTLQGLGFERGWGDTAARVSEMMHLLLDILQAPDPSTLEK
313770763 LDFLRVHKYK----------GHALMLNDRIKSVSRLQSALLKAEEYISKLPSETLYTEFEYTFQGMGFERGWGDTAARVLEMMHLLLDILQAPDPSTLET
3915037 LDFLRVHTYK----------GHALMLNDRIQSISKLQSALVKAEDHLSKLAPDTLYSEFEYELQGTGFERGWGDTAARVLEMMHLLLDILQAPDPSTLET
225437428 LDFLRVHKYK----------GQVIMLNDRIQSISRLQSALVKADDHLTKLPPETPFGEFEYEFQGMGFERGWGDTAQRVLEMIHLLLDILQAPDPSTLET
162458268 LDFLRGHRHK----------GHVMMLNDRIQSLGRLQSVLTKAEEHLSKLPADTPYSQFAYKFQEWGLEKGWGDTAGHVLEMIHLLLDIIQAPDPSTLEK
242035817 LDFLRGHRHK----------GHVMMLNDRVQSLGRLQSVLTKAEEYLSKLPAETPYAQFAYKFQEWGLEKGWGDTAEHVLEMVHLLLDIIQAPDPSTLEK
115452927 LDFLRGHRHK----------GHVMMLNDRIQSLGRLQSVLTKAEEHLSKLPADTPYSQFAYKFQEWGLEKGWGDTAGYVLEMIHLLLDVLQAPDPSTLET
241896730 LDFLREHRHK----------GHVMMLNDRIQSVGRLQSVLTKAEENLSKLPAETPYSQFANQFQEWGLEKGWGDTAEHVLEMIHLLLDILQAPDPSTLET
304651490 LDFLRGHNHK----------GNVLMLNERIQRISRLESSLNKADDYLSKLPPDTPYTDFEYALQEMGFEKGWGDTANRVLETMHLLSDILQAPDPSTLET
29289943 LDFLRGHNHK----------GNVLMLNERIQRISRLESSLNKADDYLSKLPPDTPYTEFEYALQEMGFEKGWGDTAKRVLETMHLLSDILQAPDPSTLET
115310620 LDFLRAHKHK----------GHVLMLNDRIQRISRLESALSKAEDYLAKLPQDTPYSDFEYALQELGFERGWGDTAARVLNMMHLLSDILQAPDPSTLET
4468153 LEFLRAHKHK----------GHIMMLNDRIQHIPRLESALARAEDHLSKLPPDTPYSEFEYTLQGLGFERGWGDTAERVLEMMHLLSDILQAPDASTLET
345286419 LDFLRVHRHK----------GHVLMLNDRVQRISRLESQLAKAEDYVSKLPLDTPYSEFEYALQGMGFERGWGDTAARVLEMMRLLSDVLHAPDPSTLET
4468151 LDFLRVHSHK----------GHVMMLNDKIQRVSQLEYSLAGAEEYISKVPPDTPYSELESKLQGMGFERGWGDTARRSLEMMHLLSDILQAPDPSSLES

```
151176306          LDFLRVHKHK----------GVVMMLNDRIQTIQRLQSALSKAEDYLIKLPADTPYSEFEFVIQGMGFERGWGDTAERVLEMMHLLLDILQAPDPSTLET
14530225           LDFLRVHKHK----------GVVMMLNDRIQTIQRLQSALSKAEDYLIKLPADTPYSEFEFVIQGMGFERGWGDTAERVLEMMHLLLDILQAPDPSTLET
15239816           LEFLRTHKHD----------GRPMMLNDRIQNIPILQGALARAEEFLSKLPLATPYSEFEFELQGMGFERGWGDTAQKVSEMVHLLLDILQAPDPSVLET
297795665          LEFLRTHKHD----------GRSMMLNDRIQNIPILQGALARAEEFLSKLPLATPYSEFEFELQGMGFERGWGDTAQKVSEMVHLLLDILQAPDPSVLET
225431790          LDFLRAHKHD----------GHVMMLNDRIQNISRLQSALARAEEYLSKLPPLTPYSEFEFELQGMGFEKGWGDTAQRVSEMVHLLLEILQAPDPSTLET
255551835          LEFLRTHKHD----------GHALMLNDRIQNLSSLHYALARAEEHLSKFPPNTPFSEFEFDLQSMGFERGWGDRAERVSEMVHLLMDILQAPDPASLES
168009716          FEFLRMHTYR----------GQTLMLNDRIASLVRLRPQLVKAEEALSKLPEKTPFADFAHQLQGLGLEKGWGNSAGRALETIKMLQDLLQAPDPDTLEK
168035060          FQFLRMHTYR----------GQTLMLNDRITSLRRLRPQLVKADDILSKLPEDTPFTDFAHKLQELGLEKGWGNTAGRVVETIKLLEDLLQAPDPDTLEK
168058907          FQFLRLHTYR----------GETLMLNERIATFSRFRPQLVRAEEALSKLPEDTPFSSFAHRLQELGLEKGWGNTAGRVLQTLKLLLDLLQAPDPDTLEK
168029793          VEFMRVHKYK----------DQTLLLNESITNVVRLRPALIKAEEYLIKLPNDQPLKDFYSKLQELGLERGWGDTAGRVLEMIHLLLDLLQAPDPDILEK
94266940           YEFLKLHQLH----------GIQLLLDGERVRNPVQLEDALAAALDFLERCHWPDDLERIRQRLRRLGFLDGWGDSLPRILETMHMLQDILEQPDEANLEE
94264333           YEFLKLHQLH----------GIQLLLDGERVRNPVQLEDALAAALDFLERCHWPDDLERLRQRLRRLGFLDGWGDSLPRILETMHMLQDILEQPDEANLEE
297569306          YEFLKLHQLH----------GTQLLLDGSLVRSPEELEEALSSAMDFLERCRYPDDLARISQRLGRLGFLAGWGNSLPRMLETMHMLQDILEQPDEANLEE
254416162          FNFLSLHEVQ----------GTQLLIN-GRIKSQQQLSDQVKLALGFVSDRAEDEPYESFRFKLQDMGFEAGWGNTASRVRETLEILDELIDSPDHQGLEE
332712456          YGFLKVRHFH----------GNQLLIN-ERIHNHQQLSEQVKLALEFVSDRPDSESYDKFRFKLQEMGFEPGWGNTASRVRETLAMLDELIDEPDDRALEQ
17232477           FNFLRLHNYN----------GIQLLIN-HQIQSQQQLSQQVKNALNFVSDRPNDEPYEQFRLQLQTMGFEPGWGNTASRVRDTLNILDELIDSPDPQTLEA
75908500           FNFLRLHNYN----------GIQLLIN-YQIQSQQQLSQQVKKALNFVSDRPHDEPYEQFRLQLQAMGFEPGWGNTASRVRDTLNILDELIDSPDPLTLEA
186682280          LNFLRLHHYN----------GVQLLVN-DRIQSQQQLSEQVKKAIGFVNNRPDDEPYEQFRFQLQSMGFEPGWGNTAARVRETLNILDELIDSADPQTLEA
119512682          LNFLRLHQYN----------GVQLLIN-DRITTQQQLSTQIKKAITFVSDRPKDEPYEKFRFELQMIGFEPGWGNTAKRVGDTLDILDELIDSPDPQTIEA
334117431          FNFLRLHRYN----------GVQLLIN-DRIKSQEQLSEQVKKALTFVSDLSEEEGYERFRLVLQMMGFEAGWGNTAARVHETLGILDELIDSPDPQTLEA
218440696          FNFLSLHCYN----------GITLLIN-GRIKNQHQLSEQVKDAIFFLNQFSDEIPFEDFRYEFQSMGFEPGWGNTAGRVKETLSILDELIDSPDDQVLEA
159031025          FQFLRMHQYN----------GITLLIN-DRIGNQRQLSQQVKAALDFLESYPSEEPYSNLRFELQSLGFEPGWGNTASRMRESLELLDGLIDAPDHQSLEA
220907171          FQFLSLHRYD----------GNQLLIS-ERIKNQHQLSMQVKQALHLVNSQPPQALFSDFRFELQNLGFEPGWGNTAARVKETLELLDSLIDSPDPHSLET
22298591           FNFLRIHRYN----------GYQLLIN-ERIRSPQHLSEQVKQALVVLSDRPPTEAYSEFRFELQNLGFEPGWGNTVARVRDTLEILDQLLDSPDHQVLEA
158339122          FSFLKLHSYN----------GTQLLIN-QRIQSPEHLSECVKQAISLVGGLPPEQPYPEFRFNFQELGFEPGWGNTAARVLETLEMLDELIDSPDDQVLEA
284053161          YQFLSLHSYN----------GLQLLIN-DRITNQNQLSYAIKKAISLLNKRSPKEPYEKFRFELQEIGFEPGWGNTARRALETLEILDELLDSPDHQVLEA
209523126          YQFLSLHSYN----------GLQLMIN-DRITNQNQLSHAIKKAISLLNKRSPKEPYENFRFELQEIGFEPGWGNTARRALETLEMLDELLDSPDHQVLEA
17228554           YLALRGLQYD----------GINLMIG-DAIPSGIHLAKQIHAAIKFLSALPPEEPYEKFYIELQKLGFEPGWGNTAQRILETITLLDKLIDSPQPAVLEA
75909957           YLALRGLQYD----------GINLMIG-DAIPSGIHLAKQIHAAIKFLSDLPPEEPYEKFYIELQALGFEPGWGNTAERILETITLLDRLIDSPQPAVLEA
186685043          FQALQGLQHD----------GIRLLLS-DRIPSGIHLAKQIKLALKLVNERSPHEPYEKFSLDLQELGFEPGWGNTAARVSETLELLDRLIYSPEPGILEA
119510624          FQALYQLTYD----------QKPLLIS-DRIPSGIHLVKQIKQALKFLNQQPPEEPYANFRPHLQELGFEPGWGNTSGRISETLELLEQLIDNPQPAILEA
298489784          FQALQGIQFN----------GMKVLIS-DRIHSGIQFAKLIKPAITLLSELPPDEPYAQFRSHLQELGLEAGWGNNAGRVRETLELLQRLIDTPQTYILEA
220909283          FRVLCGHEYD----------GLSLLIN-EQIGSGEDLYKQVQQALWFIRQRPGDEPWEQLHYYLEKLGFAPGWGDRASRVEETLELLQYLMDSPAPPILEA
332708740          LKNLQAHHYD----------HTPLLLN-NRIDSTTQLFEKLKEALTLVGELPAHTPYEKFRFELQVLGFEAGWGNTAGRVRETLELLERLMDAPDHAVLEA
Clustal Consensus                     ::         .         :          .   :   *:  ***         .  *   :  .    :

                                    310       320       330       340       350       360       370       380       390       400
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
114331077          FLARIPMISRLLILSPHGYFGQDNVLGLPDTGGQVVYILDQVRALEQEMRDRLQLQGVQ---VEPKILIVTRLIPDAGDTTCNQRLEKVSGCTNTWILRV
30249199           FLACIPMISRLLILSPHGYFGQDNVLGLPDTGGQVVYILDQVRALEKEMHDRLQLQGVQ---VEPKILIVTRLIPDAGDTTCNQRLEKVSGCTNTWILRV
82703384           FLARIPMISKLLILSPHGYFGQDNVLGLPDTGGQVVYILDQVRALEREMSERLILQGID---AAPKILIGTRLIPDAGDTLCHQPLEKIHGTQNSWIVRV
198283392          FIDRIPMISKILIVSIHGWFAQDKVLGRPDTGGQVVYILDQARALEQEMRQRLARQGVD---IVPRILIATRLIPNADGTTCDQRLEPVHGADNVQILRV
344199811          FIDRIPMISKILIVSIHGWFAQDKVLGRPDTGGQVVYILDQARALEQEMRQRLARQGVD---IVPRILIATRLIPNADGTTCDQRLETVHGVDNVQILRV
255021595          FLDRIPMISRILIVSIHGWFAQDKVLGRPDTGGQVVYILDQARALEREMRNRLRQQGVD---VEPRILIATRLIPESDGTTCDQRLEPVVGAENVQILRV
77166514           FLARIPMIFNIVILSPHGYFGQGNILGLPDTGGQVVYILDQVRALEKEMHRQLKEQGLD---VAPQILVVTRLIPEAQGTRCDQRLESIVGTENAAILRV
300115586          FLARIPMIFNIAILSPHGYFGQGNILGLPDTGGQVVYILDQVRALEKEMRRQLKEQGLD---VTPQILVVTRLIPEARGTRCDQRLESIVGTENAAILRV
292493898          FLARIPMIFNIVILSPHGYFGQGNILGLPDTGGQVVYILDQVRALEKEMRRQLKEEGLD---IEPQILVVTRLIPEAQGTRCDQRLEAIVGTENAAILRV
```

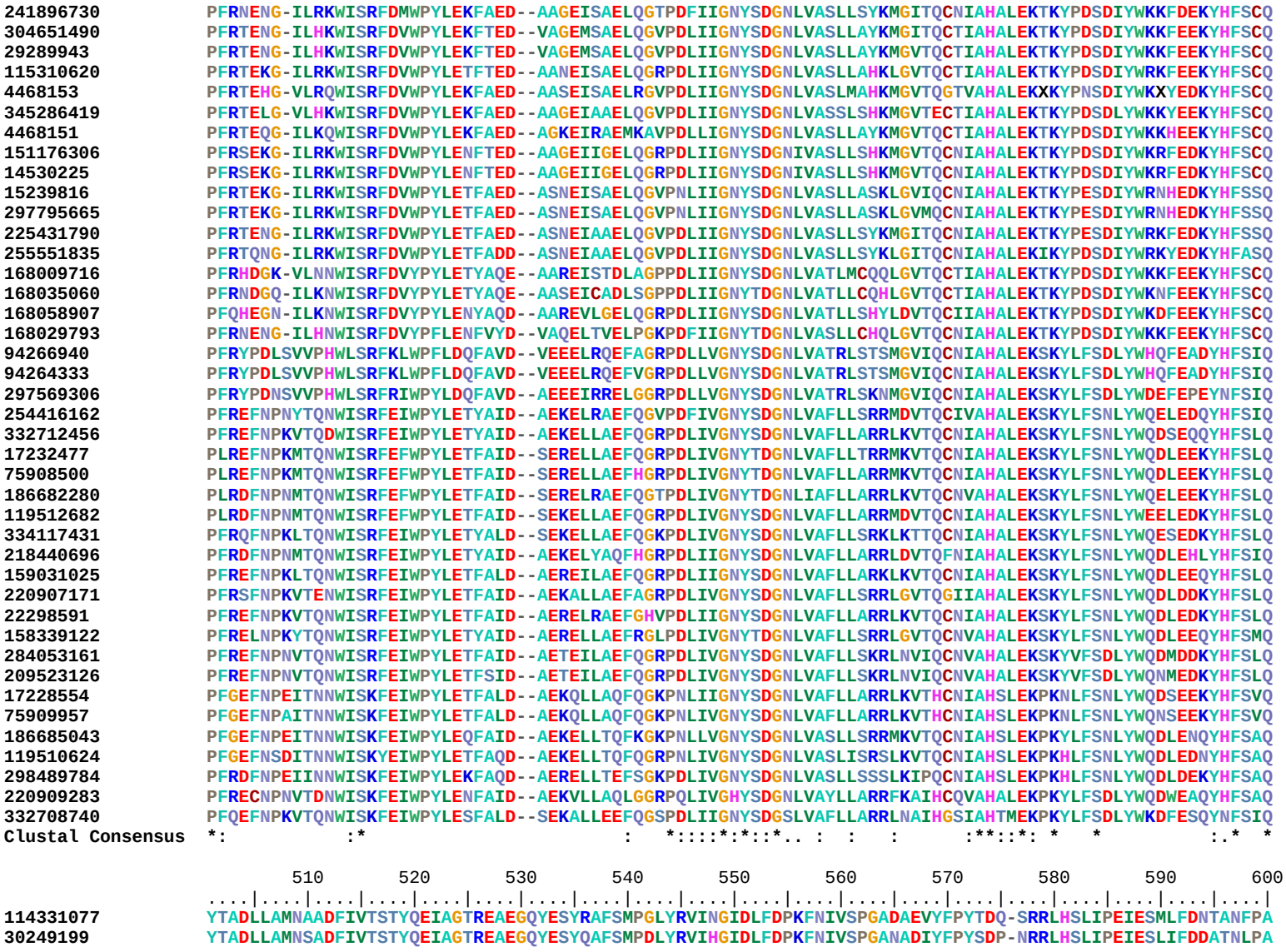

350552412 FLARIPMIFSLLILSPHGFFGQANVLGRPDTGGQVVYILDQVRALEREMRSRLQAQGLH---IEPRILVVTRLIPEAQGTSCDQPVERINGTRNAQILRV
220933364 FLARIPMIFSLLILSPHGFFGQAGVLGLPDTGGQVVYILDQVRALEREMRDRLAEQGLD---IEPRIRVVTRLIPEARGTSCDQPEEAVSGTENARILRV
350574485 FLGRIPMIFSIAILSPHGWFGQSNVLGRPDTGGQVVYILDQVRALEREMRARLAEQGID---IDPEVIVITRLIPESEGTTSDQRIEPIAGTQNARILRV
95929190 FLGRIPMIFSIVVVSPHGYFGQENVLGLPDTGGQVVYILDQVRALEKEMKEQIYRQGLD---IEPSIVVLTRLIPHCGDTSCNQPEEQIAGTSNATIVRV
298528445 FLSRIPMIFKLVIVSPHGYFGQSNVLGRPDTGGQIVYILDQVRALEKEMRRQIKEQGLE---IEPEIVVLTRQIPEAGDTTCNQRQEDIVGTSNARILRV
297602308 FFSRVPSIFNIVIFSIHGYFGQEKVLGLPDTGGQVVYILDQVRAMEEELLQRIKQQGLH---VTPKILVLTRLIPDAKGTKCNVELEPVENTKYSHILRV
115457664 FFSRVPSIFNIVIFSIHGYFGQEKVLGLPDTGGQVVYILDQVRAMEEELLQRIKQQGLH---VTPKILVLTRLIPDAKGTKCNVELEPVENTKYSHILRV
326504012 FFSRVPSIFNIVVFSIHGYFGQEKVLGLPDTGGQVVYILDQVRSMEEELLQRIKLQGLH---ITPKILVLTRLIPDSKGTKCNVELEPVENTKYSHILRV
115450038 FFSTVPCVFTVVIFSIHGYFGQEKVLGMPDTGGQVVYILDQVRALEDELLQRIKQQGLN---ATPKILVLTRLIPEAKGTKCNVELEPIENTKHSNILRV
326531526 FFSRVPSVFNIVIFSIHGYFGQEKVLGLPDTGGQVVYILDQVRALEEELLQRIKRQGLN---VTPKILVLTRLIPDAKGTKCNVELEPVEHTKHSSILRV
242063616 FFSRVPSVFDVVIFSVHGYFGQHKVLGMPDTGGQVVYILDQVRALEEELLQRIKGQGLT---FTPNILVLTRLIPEAKGTTCNVELEPIENTRHSSILRV
42568160 FFARVPRIFNVVIFSVHGYFGQTDVLGLPDTGGQVVYILDQVKALEDELLQRINSQGLN---FKPQILVVTRLIPDAKKTKCNQELEPIFGTKYSNILRI
297805240 FFARIPRIFNVVIFSVHGYFGQTDVLGLPDTGGQVVYILDQVKALEDELLHRINSQGLN---FKPQILVVTRLIPDAKNTKCNQELEPIFGTKHSNILRI
313770765 FLSRLPTVFNVVIFSPHGYFGQADVLGLPDTGGQVVYILDQVKALEEELLLRIKQQGLN---VKPQIVVATRLIPDARGTTCNLEFEAIDGTKYSNILRV
313770767 FFSRLPTVFNVVIFSPHGYFGQADVLGLPDTGGQVVYILDQVKALEDELLLRIEQQGLN---IKPQIVVVTRLIPEARGTKCNQELESINGTKHSNILRV
255584097 FLSRVPTIFNVVIFSPHGYFGQANVLGLPDTG----------------------------------GQVTRLIPDARGTKCNQELEAINGTKHSNILRV
225466221 FLSRLPTIFNVVIFSPHGYFGQSDVLGLPDTGGQVVYILDQVRALEEELLLRIKLQGLN---VKPQILVVTRLIPDARGTKCNQEWEPIDNTKHSTILRI
313770769 LFSRIPNMFNIVIFSPHGYFGQSDVLGLPDTGGQVVYILDQVRALEEELLLKIKHQGLG---VKPRILVVTRLIPNAGGTKCNQEVEPIFGTQHSHIVRV
313770771 LFSRIPNVFNIVIFSPHGYFGQSDVLGLPDTGGQIVYILDQVRALEEELLLKIRQQGLS---VKPQILVITRLIPHAGGTKCNQEVEPIFGTKHSHIVRV
255570671 LFSRLPNMFNIVIFSPHGYFGQADVLGLPDTGGQVVYILDQVRALEEELLLRIKQQGLT---MKPQILVVTRLIPDAKGTKCNQEVEPIIGTKHSNILRI
15219457 LFSRLPTVFNVVIFSVHGYFGQQDVLGLPDTGGQVVYILDQVRALEEELLIRINQQGLG---FKPQILVVTRLIPEARGTKCDQELEAIEGTKHSHILRV
297842089 LFSRLPTLFNVVIFSVHGYFGQQDVLGLPDTGGQVVYILDQVRPLEEELLIRINQQGLG---FKPQILVVTRLIPEARGTKCDQELEAIEGTKHSHILRV
115466896 FLGTIPMMFNVVILSPHGYFAQSNVLGYPDTGGQVVYILDQVRALENEMLLRIKQQGLD---ITPKILIVTRLLPDAVGTTCGQRVEKVIGTEHTDILRV
17980243 FLGTIPMTFNVVILSPHGYFAQSNVLGYPDTGGQVVYILDQVRALENEMLLRIKQQGLD---ITPKILIVTRLLPDAVGTTCGQRLEKVIGTEHTDILRV
162460741 FLGTIPMMFNVVILSPHGYFAQSNVLGYPDTGGQVVYILDQVRALENEMLLRIKQQGLD---ITPKILIVTRLLPDAAGTTCGQRLEKVIGTEHTDIIRV
239984690 FLGTIPMMFNVVILSPHGYFAQSNVLGYPDTGGQVVYILDQVRALENEMLLRIKQQGLD---ITPKILITR-LLPDAVGTTCGQRLEKVIGTEHTDILRV
162460681 FLGTIPMVFNVVILSPHGYFAQANVLGYPDTGGQVVYILDQVRAMENEMLLRIKQCGLD---ITPKILIVTRLLPDATGTTCGQRLEKVLGTEHCHILRV
242035533 FLGTIPMVFNVVILSPHGYFAQANVLGYPDTGGQVVYILDQVRAMENEMLLRIKQCGLD---ITPKILIVTRLLPDATGTTCGQRLEKVLGTEHCHILRV
62865493 FLGTIPMVFNVVILSPHGYFAQANVLGYPDTGGQVVYILDQVRAMENEMLLRIKQQGLN---ITPRILIVTRLLPDATGTTCGQRLEKVLGTEHTHILRV
17980241 FLGTIPMVFNVVILSPHGYFAQANVLGYPDTGGQVVYILDQVRAMENEMLLRIKQQGLN---ITPRILIVTRLLPDATGTTCGQRLEKVLGTEHTHILRV
115453437 FLGTIPMVFNVVIMSPHGYFAQANVLGYPDTGGQVVYILDQVRAMENEMLLRIKQQGLN---ITPRILIVTRLLPDATGTTCGQRLEKVLGTEHTHILRV
115473359 FLGTIPMVFNVVILSPHGYFAQANVLGYPDTGGQVVYILDQVRAMENEMLLRIKQQGLN---ITPRILIVTRLLPDAHGTTCGQRLEKVLGTEHTHILRV
3980298 FLGTIPMVFNVVILSPHGYFAQANVLGYPDTGGQVVYILDQVRAMENEMLLRIKQQGLD---ITPKILIVTRMLPDAHGTTCGQRLEKVLGTEHTHILRV
15242073 FLGRVPMVFNVVILSPHGYFAQDNVLGYPDTGGQVVYILDQVRALEIEMLQRIKQQGLN---IKPRILILTRLLPDAVGTTCGERLERVYDSEYCDILRV
297812265 FLGRVPMVFNVVILSPHGYFAQDNVLGYPDTGGQVVYILDQVRALETEMLQRIKQQGLD---IKPRILILTRLLPDAVGTTCGERLERVYDSEYCDILRV
22331535 FLGRIPMVFNVVILSPHGYFAQDNVLGYPDTGGQVVYILDQVRALETEMLQRIKQQGLN---ITPRILIITRLLPDAAGTTCGQRLEKVYGSQYCDILRV
297818772 FLGRIPMVFNVVILSPHGYFAQDNVLGYPDTGGQVVYILDQVRALETEMLQRTKQQGLT---ITPRILIITRLLPDAAGTTCGQRLEKVYGSQYCDILRV
313770759 FLGRIPMVFNVVIMSPHGYFAQDNVLGYPDTGGQVVYILDQVRALENEMLQRIKKQGLD---IIPRILIITRLLPDAVGTTCGQRLERVYGSEHCDILRV
313770761 FLGRIPMVFNVVIMSPHGYFAQDNVLGYPDTGGQVVYILDQVRALESEMLLRIKQQGLD---ITPRILIITRLLPDAVGTTCGQRLEKVYGSEHCDILRV
255550319 FLGRIPMVFNVVIMSPHGYFAQDNVLGYPDTG----------------------------GQITRLLPDAVGTTCGQRLEKVFGTEHSDILRI
324984223 FLGRIPMVFNVVILTPHGYFAQDNVLGYPDTGGQVVYILDQVRALENEMLLRIKQQGLN---ITPRILIITRLLPDAVGTTCGQRLEKVYGTEYSDILRV
258489633 FLGRIPMVFNVVILTPHGYFAQDNVLGYPDTGGQVVYILDQVRALENEMLLRIKQQGLN---ITPRILIITRLLPDAVGTTCGQRLEKVYGTEYSDILRV
324984221 FLGRIPMVFNVVILTPHGYFAQDNVLGYPDTGGQVVYILDQVRALENEMLLRIKQQGLN---ITPRILIITRLLPDAVGTTCGQRLEKVYGTEHSDILRV
324984229 FLGRIPMVFNVVILTPHGYFAQDNVLGYPDTGGQVVYILDQVRALENETLLRIKQQGLN---ITPRILIITRLLPDAVGTTCGQRLEKVYGTEHSDILRV
6683114 FLGRIPMVFNVVILTPHGYFAQDDVLGYPDTGGQVIYILDQVRALEDEMLLRIKQQGLD---ITPQILIITRLLPDAVGTTCGQRLEKVYGTKYSDILRV
4584690 FLDRIPMVFNVVILSPHGYFAQDDVLGYPDTGGQVVYILDQVRALESEMLSRIKKQGLD---IIPRILIITRLLPDAVGTTCGQRLEKVYGTEHCHILRV
3915046 FLDRIPMVFNVVILSPHGYFAQDDVLGYPDTGGQVVYILDQVRALESEMLSRIKKQGLD---IIPRILIITRLLPDAVGTTCGQRLEKVYGTEHCHILRV


3766299 FLDRIPMVFNVVILSPHGYFAQDDVLGYPDTGGQVVYILDQVRALESEMLNRIKKQGLD---IVPRILIITRLLPDAVGTTCGQRLEKVYGTEHCHILRV
3377764 FLDRIPMVFNVVILSPHGYFAQDDVLGYPDTGGQVVYILDQVRALESEMLNRIKKQGLD---IVPRILIITRLLPDAVGTTCGQRLEKVYGTEHCHILRV
225444613 FLGRIPMVFNVVILSPHGYFAQDNVLGYPDTGGQVVYILDQVRAMETEMLLRIKQQGLD---ITPKIIIVTRLLPDAVGTTCNQRIEKVYGTEHSIILRV
3660531 FLGRIPMVFNVVILSPHGYLAQENVLGYPDTGGQVVYILDQVPALEREMLKRIKEQGLD---IIPRILIVTRLLPDAVGTTCGQRLEKVYGTEHSHILRV
31455440 FLGRIPMVFNVVILSPHGYFAQENVLGYPDTGGQVVYILDQVPALEREMLKRIKEQGLD---IIPRILIVTRLLPDAVGTTCGQRIEKVYGAEHSHILRV
304651488 FLSRIPMVFNVVILSPHGYFAQENVLGYPDTGGQVVYILDQVPALEREMLKRIKEQGLD---IKPRILIVTRLLPDAVGTTCGQRLEKVFGTEHSHILRV
28629438 FLSRIPMVFNVVILSPHGYFAQENVLGYPDTGGQVVYILDQVPALEREMLKRIKEQGLD---IKPRILIVTRLLPDAVGTTCGQRLEKVFGTEHSHILRV
115310618 FLGRIPMVFNVVILSPHGYFAQENVLGYPDTGGQVVYILDQVPALEREMLKRIKEQGLD---VKPRILIITRLLPDAPGTTCGQRLEKVYGSEYSHILRV
345286417 FLGRIPMVFNVVILSPHGYFAQENVLGYPDTGGQVVYILDQVPALEREMLKRIKEQGLD---ITPRILIVTRLLPDAVGTTCGQRLEKVFGAEHSHILRV
38425095 FLGRIPMVFNVVILTPHGYFAQANVLGYPDTGGQVVYILDQVRALEHEMLLRIKQQGLD---IIPRILIVSRLLPDAVGTTCGQRLEKVFGTEHSHILRV
15235300 FLGMVPMVFNVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALETEMLLRIKRQGLD---ISPSILIVTRLIPDAKGTTCNQRLERVSGTEHTHILRV
297814081 FLGMVPMVFDVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALESEMLLRIKRQGLD---ITPRILIVTRLIPDAKGTTCNQRLERVSGTEHTHILRV
6682841 FLGRLPMVFNVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALENEMLLRIKRQGLD---ISPKILIVTRLIPDAKGTTCNQRLERVSGTEHTHILRV
255564236 FLGRIPMVFNVVILSPHGYFGQANVLGLPDTG----------------------------------GQVTRLIPDAKGTTCNQRLERVSGTEYTHILRV
313770763 FLGRVPMVFNVVILSPHGYFGQANVLGLPDTGGQIVYILDQVRALENEMLLRIQQQGLD---FKPKILIVTRLIPDSKGTSCNQRLERVSGTEHTHILRV
3915037 FLGRVPMVFNVVILSPHGFFGQANVLGLPDTGGQVVYILDQVRALESEMLVRIKKQGLD---FTPRILIVTRLIPDAKGTTCNQRLERVSGTEYTHILRV
225437428 FLGRIPMVFNVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALENEMLLRMQKQGLD---VTPRILIVTRLIPDAKGTTCNQRLERVSGTEHSHILRV
162458268 FLGRIPMIFNVVVVSPHGYFGQANVLGLPDTGGQIVYILDQVRALENEMVLRLKKQGLD---VSPKILIVTRLIPDAKGTSCNQRLERISGTQHTYILRV
242035817 FLGRIPMIFNVVVVSPHGYFGQANVLGLPDTGGQIVYILDQVRALENEMVLRLKKQGLD---FSPKILIVTRLIPDAKGTSCNQRLERISGTQHTYILRV
115452927 FLGRIPMIFNVVVVSPHGYFGQANVLGLPDTGGQIVYILDQVRALENEMVLRLKKQGLD---FTPKILIVTRLIPEAKGTSCNQRLERISGTQHTYILRV
241896730 FLGRIPMIFNVVIVSPHGYFGQANVLGMPDTGGQIVYILDQVRALENEMVLRLKKQGLD---VTPKILIVTRLIPDSKGTSCNQRLERISGTQHTYILRV
304651490 FLGRLPMVFNVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALEAEMLLRIKQQGLN---FKPRILVVTRLIPDAKGTTCNQRLERISGTEYSHILRV
29289943 FLGRLPMVFNVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALEAEMLLRIKQQGLN---FKPKILVVTRLIPDAKGTTCNQRLERISGTEYSHILRV
115310620 FLGRVPMVFNVAILSVHGYFGQANVLGLPDTGGQIVYILDQVRALENEMLLRIKQQGLN---VTPRILIVTRLIPDAKGTTCNQRLERVSGTEYTSILRV
4468153 FLGRIPMVFNVVILSIHGYFGQANVLGLPDTGGQIVYILDQVRALENEMIKRIKAQGLS---IIPQILIVTRLIPDAKGTSCNQRLEKISGCEHSHILRV
345286419 FLGRVPMVFNVVILSVHGYFGQANVLGLPDTGGQIVYILDQVRALESETLQRIKKQGLQ---ITPRILVVTRLIPDAADTSCNQRLERLSGCEYSHILRV
4468151 FLGQLPMVFNVVVLSIHGYFAQTDVLGLPDTGGQVVYVLDQVRAMENEMIKRIKNHGLN---ITPRILIVTRLIPEARGTKCDQRLEKIDGCEHSHILRV
151176306 FLGRLPMVFNVVILSVHGYFGQAHVLGLPDTGGQIVYILDQVRSLEHEMLQRIKKQGLD---VTPRILIVSRLIPDAKGTTCNQRMEKVSGTEHASILRV
14530225 FLGRLPMVFNVVILSVHGYFGQAHVLGLPDTGGQIVYILDQVRSLEHEMLQRIKKQGLD---VTPRILIVSRLIPDAKGTTCNQRMEKVSGTEHASILRV
15239816 FLGRIPMVFNVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALENEMLLRIQKQGLE---VIPKILIVTRLLPEAKGTTCNQRLERVSGTEHAHILRI
297795665 FLGRIPMVFNVVILSPHGYFGQANVLGLPDTGGQVVYILDQVRALENEMLLRIQKQGLE---VIPKILIVTRLIPEAKGTTCNQRLEKVSGTEHAHILRI
225431790 FLGRIPMVFNVVIVSPHGYFGQANVLGLPDTGGQIVYILDQVRALENEMLLRIQKQGLD---VIPKILIVTRLIPDAKGTTCNQRLERISGTEHTHILRV
255551835 FLGMLPMVFNVVIVSPHGYFGQANVLGLPDTG----------------------------------GQVTRLIPHAKGTTCNQRLERISGTENTYILRV
168009716 FLARIPMVFSVVIVSPHGYFGQEGVLGLPDTGGQVVYILDQVRALENEMLENLQLQGLD---IIPQIVILTRLIPNAIGTTCNQRIEKVTGSRFSHILRI
168035060 FLARIPMVFSVVIVTPHGYFGQDGVLGLPDTGGQVVYILDQVRALENEMLENLQLQGLD---IVPKIVILTRLIPNAFGTTCNQRIEKVHGSRFSHILRI
168058907 FLARIPMIFTVCIVSPHGYFGQAGVLGLPDTGGQVVYILDQVRALENQMLENLQLQGLD---FKP--QILTRLIPNANGTTVNQRIEKVSGTQHSRILRV
168029793 FLARIPIVFSVAIISPHGYFGQSNVLGMPDTGGQVVYILDQVRAMEKEMLKNIKLQGLD---IEPQIVVVTRLIPNANGTTCNQRIEQIEGTKHSRILRV
94266940 FLARIPMVSKVALISPHGWFGQDNVLGRPDTGGQVVYILDQARALEQFLAADLRSAG---LEIEPKILIITRLIPENEGTTADQRLEPVRDTANVAILRV
94264333 FLSRIPMVSKVALISPHGWFGQDNVLGRPDTGGQVVYILDQARALEQFLAADLRSAG---LEIEPKILIITRLIPENEGTTADQRLEPVRDTANVAILRV
297569306 FLSRIPMVSSVALISPHGWFGQENVLGRPDTGGQVVYVLDQAKALEDFLAQDLRDAG---LEIAPKILIVSRLIPENEGTSADQRLEKVYDTDDVWILRV
254416162 FLSRIPMIFRIVLVSIHGWFGQEGVLGRPDTGGQVVYVLDQARSLEKQLQEDIKLAGLELLKVKPKVIILSRLIPNNDGTRCNERLEKVHGTDNAWILRV
332712456 FLSRIPMIFRIVLVSVHGWFGQEGVLGRPDTGGQVVYVLDQAKSLEKQLQENLTLAG---LNIQPKVIILTRLIPNNDGTRCNERLEKIKGTENAWILRV
17232477 FISRIPMIFRIVLVSAHGWFGQEGVLGRPDTGGQVVYVLDQAKNLEKQLQEDAILAGLEVLNVQPKVIILTRLIPNSDGTLCNQRLEKVYGTENAWILRV
75908500 FISRIPMIFRIVLVSAHGWFGQEGVLGRPDTGGQVVYVLDQAKNLEKQLQEDAILAGLEVLNVQPKVIILTRLIPNSDGTLCNQRLEKVHGTENAWILRV
186682280 FISRVPMIFRIVLVSAHGWFGQEGVLGRPDTGGQVVYVLDQAKSLEKQLQEDVLLAGLEKLNVEPKVIILTRLIPNSDGTLCNQRLEKVHGTENAWILRV
119512682 FISRIPMIFRIVLVSAHGWFGQEGVLGRPDTGGQVVYVLDQAKNLEKQLQEDAQLAGLDGLNVEPKVIILTRLIPNSDGTLCNQRLEKVHGTENAWILRV
334117431 FISRIPMIFKIVLVSPHGWFGQEGVLGRPDTGGQVVYVLDQAKSLEKQLQEDIHLAGLDSLGVKPKVIILTRLIPNSDGTRCNERLEKVHATENAWILRV

```
218440696          FLSRVPMIFRIVLVSVHGWFGQEGVLGRPDTGGQVVYVLDQARSLEQQLKEDIILAGLDGYGIEPKVIILSRLIHNSDGTRCNQRLEKVHGTDNAWILRV
159031025          FLSRIPMLFRIVLVSVHGWFGQEGVLGRPDTGGQVVYVLDQARSLEQQLREDIFLAGLEGLGIEPKIIILTRLLPNSEGTRCDQRLEKVYGTDNACILRV
220907171          FLSRIPMIFRIALISPHGWFGQEGVLGRPDTGGQVVYVLDQARSLEQQLREEITLAGLEGLNVHPKVIILTRLIPNSDGTRCNQPLEKVYGTENVWILRV
22298591           FVSRIPMLFRIALISPHGWFGQEGVLGRPDTGGQVVYILDQVKSLEKQMREDLELAGLGVLEAQPKIIVLTRLIPNAEGTLCNQRLEKIYGTNDAWILRV
158339122          FISRIPMIFKIVLVSIHGYFGQEGVLGRPDTGGQVVYVLDQAKSLEKQLQEDLQFAGLDTLEVQPKLIILSRLIPNSEGTLCNQRLEKVHATDNVWILRV
284053161          FLARIPMIFRIVLVSIHGWFAQEGVLGRPDTGGQIVYILDQARSLEMELEEELKLSGLSVLGVQPKVMILTRLIPHSDGTRCDQRLEKVYGTKNAWILRV
209523126          FLARIPMIFRIVLVSIHGWFGQEGVLGRPDTGGQVVYILDQARSLEMELEEELKLSGLSVLGVQPKVMILTRLIPHSDGTRCHQRLEKVYGTKNAWILRV
17228554           FVARVPAVFRVVLVSIHGWVAQEDVMGRDETLGQVIYVLEQARSLENKLQQEIKLAGLEVLGIQPHVIILTRLIPNCEGTYCNLRLEKLHNTENAWILRV
75909957           FVARVPAVFRVVLVSIHGWVAQEDVMGRDETLGQVIYVLEQARSLENKLQQEIKLAGLEVLGIQPHIIILTRLIPHCEGTYCNLRLEKLHNTENAWILRV
186685043          FVARVPAVFRVVLISIHGWVGQEDVVGRDETLSQVIYVLEQARSLENELREQIKLAGLDQLGIKPHVIILTRLIPNCEGTFCYLPLEKVQDTENAWILRV
119510624          FVARVPAIFRVVLVSIHGWVGQQDVLGRDETLGQVIYVLEQARSLENKLQAEIKLAGLDLIGIQPHVIILTRLIPNCEGTECNLRLEKIHDTENAWILRI
298489784          FVARIPAVFRVVLISIHGWVAQEDVLGRDETLGQVIYVLEQARSLENKLQQEIKLAGLDFLGIKPHVIILTRLIPNCESTFCDLRLEKVHNTENAWILRV
220909283          FLARIPIVFRAVLVSIHGWVGQSGVMGRSETMSQVSYVLEQARQLEHELQSDIKRSGLDRLGIRPQVVILTRLIPHCEETLCYLPLEKLAETMNGWILRV
332708740          FISRIPLIFRVVLVSVHGWVGQEGVLGLPDTAGQVAYVIDQARSLEQTIQNNIKLSGLDVLGVEPKVIVLTRLIPNCEGTQCNLRLEKIQGTSNGWILRV
Clustal Consensus  :.  :*      :.: **:..*  ::*  :*                                          : .   *      * :       *:*:

                                 410       420       430       440       450       460       470       480       490       500
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
114331077          PFRKKNGEIIPQWISRFEIWPHLETFALD--VEREALAELGRRPDLIIGNYSDGNLVATLLSRRLGVTQCNIAHALEKTKYLHSDIYWQENEDKYHFSCQ
30249199           PFRKHNGEIIPHWISRFEIWPHLEIFAGD--VEREALAELGGHPDLIIGNYSDGNLVATLLSRRLGVTQCNIAHALEKTKYLHSDIYWQENEDKYHFSCQ
82703384           PFRKGSGEIVRHWISRFEIWPYLENFAHD--IEREALAQLSGRPDLIIGNYSDGNLVASLISKRIGVTQCNIAHALEQSKYLHSALYWRENEAQYHFNCQ
198283392          PFRYANGEILPQWISRFNVWPWLERYADD--LERETLAEFGRRPDLIIGNYSDGNLVATILSARLNVTQCNIAHALEKSKYLYSDLYWRDHDASHHFACQ
344199811          PFRYPNGEILPQWISRFNVWPWLERYADD--LERETLAEFGRRPDLIIGNYSDGNLVASMLSERLNVTQCNIAHALEKSKYLYSDLYWRDHDASHHFACQ
255021595          PFRYPDGRIHPHWISRFKIWPWLERYAQD--LEREVLAELGSRPDLIIGNYSDGNLVATLLSERLGVTQCNIAHALEKSKYLYSDLHWRDHEQDHHFACQ
77166514           PFRNAGGEVLPYWLSRFEVWPYLERYAMD--AEREMLAELEGSPDLIIGNYSDGSLVATLLSQRLRVTQCNIAHALEKAKYLYSDLYWRENDAQYHFACQ
300115586          PFRNAAGEVLPYWLSRFEVWPYLERYAMD--VEREMLAELEGSPDLIIGNYSDGSLVATLLSQRLRVTQCNIAHALEKTKYLYSDLYWRENDAQYHFACQ
292493898          PFRSAAGEALPYWLSRFEVWPYLERYAMD--VEREILAELEGSPDLIIGNYSDGNLVATLLAHRLRVTQCNIAHALEKTKYLYSDLYWRENDAQYHFSCQ
350552412          PFRSKEGEIIPHWISRFEIWPYLERFADD--VEQTVRAELGTRPDLIIGNYSDGNLVATLLSARMQVTQCHIAHALEKTKYLYSDLYWKENEAQYHFSCQ
220933364          PFRREDGEVVPQWISRFEIWPYLERFADE--VERTILADLGGRPDLIIGNYSDGNLVASLLSARLHVTQCNIAHALEKTKYLYSDLYWKDNDAQYHFATQ
350574485          PFRNENGDILPHWISRFHIWPYLERFALD--AETELLAELGDRPDLIIGNYSDGNLVASLMSRRLGVSQCNIAHALEKTKYLFSDLYWRDNEDRYHFSCQ
95929190           PFRNDQGEVINDWISRFKIWPHLERFSRE--SERKLLETIGARPDLIIGNYSDGNLVSFLLSRRLRVTQCTIAHALEKAKYLFSGLYWKEN-PEYNFQTQ
298528445          PFRYPSGEIVPHWISRFHVWPFLERFALD--STQEVHSELKGRPDLIIGNYSDGNLVASLMSKKMNITQCNIAHALEKSKYLFSSQYWKDNEAQYRFSSQ
297602308          PFKTEDGKDLRQWVSRFDIYPYLERYAQN--SCAKILDILEGKPDLIIGNYTDGNLVASLLSNKLCVTQGTIAHALEKTKYEDSDVKWREMDQKYHFSCQ
115457664          PFKTEDGKDLRQWVSRFDIYPYLERYAQD--SCAKILDILEGKPDLIIGNYTDGNLVASLLSNKLCVTQGTIAHALEKTKYEDSDVKWREMDQKYHFSCQ
326504012          PFKTEDGKDLRQWVSRFDIYPYLERYAQD--ASTKILDMLEGKPDLIIGNYTDGNLVASLMSSKLGVTQGTIAHALEKTKYEDSDVKWRELDQKYHFSCQ
115450038          PFKTEDGKVLPQWVSRFDIYPYLERYAQD--SSVKILEILEGKPDLVIGNYTDGNLVASLLTSKLGVTQGTIAHALEKTKYEDSDIKWRELDHKYHFSCQ
326531526          PFKTDDGKDLRQWVSRFDIYPYLERYAKD--SSVKILDILEGKPDMVIGNYTDGNLVASLLSSKLGVTQGTIAHALEKTKYEDSDVKWREMDHKYHFSCQ
242063616          PFKTQDGQDLPHWVSRFDIYPYLERYAQ-------ILDILGRKPDLVIGNYTDGNLVAYLVSRKLGVTQGTIAHALEKTKYEDSDVKWREMDHKYHFSCQ
42568160           PFVTENG-ILRRWVSRFDIYPYLERFTKD--ATTKILDILEGKPDLIIGNYTDGNLVASLMANKLGITQATIAHALEKTKYEDSDIKWKEFDPKYHFSSQ
297805240          PFVTESG-ILRRWVSRFDIYPYLERFTKD--ATTKILDILEGKPDLIIGNYTDGNLVASLMANTLGITQATIAHALEKTKYEDSDIKWKEFDPKYHFSSQ
313770765          PFRVENR-VLRQWVS----------RFDE--VTTKILDLMEGKPDLIIGNYTDGNFAATLMAGKLGITQATIAHALEKTKYENSDVKWKELESKYHFPCQ
313770767          PFSIENK-VLRQWVS----------RFDD--VITKLLDLMQRKPDLIIGNYTDGNLAATLMASKLGITQATIAHALEKTKYENSDVKWKELDPKYHFSCQ
255584097          PFTVENR-VLRQWVSRFDIYPYLEKFTQD--VADKILDLMDGKPDLIIGNYTDGNLAATLLANKLGITQATIAHALEKTKYEDSDIKWKELDPKYHFSCQ
225466221          PFRTEKG-ILNQWVSRFDIYPYLERFTQD--ATAKIIEHMEGKPDLIIGNYTDGNLVASLMATKLGITQGTIAHALEKTKYEDSDVKWKELEPKYHFSCQ
313770769          PFKTEKG-VLPQWVS----------RFDD--AADKVLEHMDSKPDLIIGNYSDGNLVASLMARKLSITLGTIAHALEKTKYEDSDVKWKELDAKYHFSCQ
313770771          PFKTEKG-VLPQWVSRFDVYPYLERFAQD--AADKVREHMDCKPDLLIGNYSDGNLVASLMAQKLGTTLGTIAHALEKTKYEDSDAKWKELDPKYHFSCQ
255570671          PFKTEKG-VLPQWVSRFDIYPYLEKFAQD--AADKVLEHMECKPDLIIGNYSDGNLVATLMANRLGITLGTIAHALEKTKYEDSDAKWKQLDPKYHFSCQ
```

```
15219457        PFVTNKG-VLRQWVSRFDIYPYLERFTQD--ATSKILQRLDCKPDLIIGNYTDGNLVASLMATKLGVTQGTIAHALEKTKYEDSDAKWKELDPKYHFSCQ
297842089       PFVTDKG-ILRQWVSRFDIYPYLERFTQD--ATSKILQRLDCKPDLIIGNYTDGNLVASLMATKIGVTQGTIAHALEKTKYEDSDAKWKELDPKYHFSCQ
115466896       PFRSENG-ILRKWISRFDVWPFLETYTED--VANEIMREMQAKPDLIIGNYSDGNLVATLLAHKLGVTQCTIAHALEKTKYPNSDIYLDKFDSQYHFSCQ
17980243        PFRTENG-ILRKWISRFDVWPFLETYTED--VANEIMREMQAKPDLIIGNYSDGNLVATLLAHKLGVTQCTIAHALEKTKYPNSDIYLDKFDSQYHFSCQ
162460741       PFRNENG-ILRKWISRFDVWPYLETYTED--VSSEIMKEMQAKPDLIIGNYSDGNLVATLLAHKLGVTQCTIAHALEKTKYPNSDIYLDKFDSQYHFSCQ
239984690       PFRTENG-IR-KWISRFDVWPYLETYTEVNDVANELMREMQTKPDLIIGNYSDGNLVATLLAHKLGVTQCTIAHALEKTKYPNSDIYLDKFDSQYHFSCQ
162460681       PFRTENG-IVRKWISRFEVWPYLETYTDD--VAHEIAGELQANPDLIIGNYSDGNLVACLLAHKMGVTHCTIAHALEKTKYPNSDLYWKKFEDHYHFSCQ
242035533       PFRTENG-IVRKWISRFEVWPYLETYTDD--VAHEIAGELQANPDLIIGNYSDGNLVACLLAHKMGVTHCTIAHALEKTKYPNSDLYWKKFEDHYHFSCQ
62865493        PFRTENG-IVRKWISRFEVWPYLETFTDD--VAHEIAGELQANPDLIIGNYSDGNLVACLLAHKMGVTHCTIAHALEKTKYPNSDLYWKKFEDHYHFSCQ
17980241        PFRTENG-IVRKWISRFEVWPYLETFTDD--VAHEIAGELQANPDLIIGNYSDGNLVACLLAHKMGVTHCTIAHALEKTKYPNSDLYWKKFEDHYHFSCQ
115453437       PFRTENG-IVRKWISRFEVWPYLETFTDD--VAHEIAGELQANPDLIIGNYSDGNLVACLLAHKMGVTHCTIAHALEKTKYPNSDLYWKKFEDHYHFSCQ
115473359       PFRTENG-TVRKWISRFEVWPYLETYTDD--VAHEISGELQATPDLIIGNYSDGNLVACLLAHKLGVTHCTIAHALEKTKYPNSDLYWKKFEDHYHFSCQ
3980298         PFKTEDG-IVRKWISRFEVWPYLEAYTDD--VAHEIAGELQANPDLIIGNYSDGNLVACLLAHKLGVTHGTIAHALEKTKYPNSDLYWKKFEDHYHFSCQ
15242073        PFRTEKG-IVRKWISRFEVWPYLETYTED--AAVELSKELNGKPDLIIGNYSDGNLVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLDDKYHFSCQ
297812265       PFRTEKG-IVRKWISRFEVWPYLETYTED--AAVELSKELNGKPDLIIGNYSDGNLVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLDEKYHFSCQ
22331535        PFRTEKG-IVRKWISRFEVWPYLETFTED--VAAEISKELQGKPDLIIGNYSDGNLVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLDEKYHFSCQ
297818772       PFRTEKG-IVRKWISRFEVWPYLETFTED--VAAEISKELQGKPDLIIGNYSDGNLVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLDEKYHFSCQ
313770759       PFRDGKG-MVRKWISRFEVWPYLETFTED--VAAEIAKELQGKPDLIIGNYSDGNIVASLLAHKLGVTECTIAHALEKTKYPDSDIYWKKFDEKYHFSCQ
313770761       PFRDEKG-MVRKWISRFEVWPYLETYTED--VAAEIAKELQGKPDLIIGNYSDGNVVASLLAHKLGVTECTIAHALEKTKYPDSDIYWKKFDEKYHFSCQ
255550319       PFRTEKG-IVRKWISRFEVWPYLETYTED--VATEIGKEFQGKPDLIIGNYSDGNIVASLLAHKLGVTECTIAHALEKTKYPESDIYWKKLDDKYHFSCQ
324984223       PFRTEKG-IVRKWISRFEVWPYLETYTED--VAHEISKELQGKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLEDKYHFSCQ
258489633       PFRTEKG-IVRKWISRFEVWPYLETYTED--VAHEISKELQGKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLEDKYHFSCQ
324984221       PFRTEKG-IVRKWISRFEVWPYLETYTED--VAHEISKELQGKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLEDKYHFSCQ
324984229       PFRTEKG-IVRKWISRFEVWPYLETYTED--VAHEISKELQGKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKLEDKYHFSCQ
6683114         PFRTEKG-VVRKWISRFEVWPYLETYTED--VAVEIAKELQGKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKNLDDKYHFSCQ
4584690         PFRDTKG-IVRKWISRFEVWPYLETYTED--VAHELAKELQGKPDLIVGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPESDIYWKKFEEKYHFSCQ
3915046         PFRDEKG-IVRKWISRFEVWPYLETYTED--VAHELAKELQSKPDLIVGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPESDIYWKKFEEKYHFSCQ
3766299         PFRDQKG-IVRKWISRFEVWPYLETYTED--VAHELAKELQGKPDLIVGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPESDIYWKKFEEKYHFSCQ
3377764         PFRDQKG-IVRKWISRFEVWPYLETYTED--VAHELAKELQGKPDLIVGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPESDIYWKKFEEKYHFSCQ
225444613       PFRTEKG-IVRKWISRFEVWPYLETYTED--VAKELATELQTKPDFIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPESDIYWKKLEDKYHFSCQ
3660531         PFRTEKG-IVRKWISRFEVWPYMETFIED--VAKEISAELQAKPDLIIGNYSEGNLAASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKFDEKYHFSSQ
31455440        PFRTEKG-IVRKWISRFEVWPYMETFIED--VAKEISAELQAKPDLIIGNYSEGNLAASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKKFDEKYHFSSQ
304651488       PFRTEKG-IVRKWISRFEVWPYMETFIED--VGKEITAELQAKPDLIIGNYSEGNLAASLLAHKLGVTQCTIAHALEKTKYPDSDIYLNKFDEKYHFSAQ
28629438        PFRTEKG-IVRKWISRFEVWPYMETFIED--VGKEITAELQAKPDLIIGNYSEGNLAASLLAHKLGVTQCTIAHALEKTKYPDSDIYLNKFDEKYHFSAQ
115310618       PFRTEKG-VVRKWISRFEVWPYMETFTED--VAKEVTAELQAKPDLVIGNYSEGNLVASLLAHKLGVTQCTIAHALEKTKYPDSDIYLSKFDEKYHFSCQ
345286417       PFRTEKG-ILRKRISRFEVWPYMETFTED--VAKEITAELQSKPDLIIGNYSEGNLAASLLAHKLGVTQCTIAHALEKTKYPDSDIYLKNFDDKYHFSCQ
38425095        PFRTEKG-IVRRWISRFEVWPYLETYTED--VANEIAGELQAKPDLIIGNYSDGNIVASLLAHKLGVTQCTIAHALEKTKYPNSDIYWKSFEEKYHFSCQ
15235300        PFRSEKG-ILRKWISRFDVWPYLENYAQD--AASEIVGELQGVPDFIIGNYSDGNLVASLMAHRMGVTQCTIAHALEKTKYPDSDIYWKDFDNKYHFSCQ
297814081       PFRSDKG-ILHKWISRFDVWPYLENYAQD--AASEIVGELQGVPDFIIGNYSDGNLVASLMAHRMGVTQCTIAHALEKTKYPDSDIYWKDFDNKYHFSCQ
6682841         PFRSEKG-ILRQWISRFDVWPYLETFTED--VGSEITAELQGFPDFIIGNYSDGNLVASLLAYKMGITQCTIAHALEKTKYPDSDIYWKKFDEKYHFSCQ
255564236       PFRSEKG-ILRKWISRFDVWPYLETLLS------EIVAELQGIPDFIIGNYSDGNLVASLLAYKMGVTQCTIAHALEKTKYPDSDIYWKKFDDKYHFSCQ
313770763       PFRSEHG-ILRKWISRFDVWPYLETFAED--AASEIVAELQGIPDFIIGNYSDGNLVASLLAYKMGVTQCTIAHALEKTKYPDSDIYWKKFDDKYHFSCQ
3915037         PFRSEKG-ILRKWISRFDVWPFLETFAED--VASEIAAELQCYPDFIIGNYSDGNLVASLLAYKMGVTQCTIAHALEKTKYPDSDIYWKKFEDKYHFSCQ
225437428       PFRTDKG-ILRKWISRFDVWPYLETFAED--AASEIAAELQGVPELIIGNYSDGNLVASLLASKLGVTQCTIAHALEKTKYPDSDIYWKNFDDKYHFSCQ
162458268       PFRNENG-ILKKWISRFDVWPYLETFAED--AAGEIAAELQGTPDFIIGNYSDGNLVASLLSYKMGITQCNIAHALEKTKYPDSDIFWKNFDEKYHFSCQ
242035817       PFRNENG-ILKKWISRFDVWPYLETFAED--AAGEIAAELQGTPDFIIGNYSDGNLVASLLSYKMGITQCNIAHALEKTKYPDSDIYWKKFDEKYHFSCQ
115452927       PFRNENG-ILRKWISRFDVWPYLEKFAED--AAGEIAAELQGTPDFIIGNYSDGNLVASLLSYKMGITQCNIAHALEKTKYPDSDIYWTKYDEKYHFSCQ
```


241896730 PFRNENG-ILRKWISRFDMWPYLEKFAED--AAGEISAELQGTPDFIIGNYSDGNLVASLLSYKMGITQCNIAHALEKTKYPDSDIYWKKFDEKYHFSCQ
304651490 PFRTENG-ILHKWISRFDVWPYLEKFTED--VAGEMSAELQGVPDLIIGNYSDGNLVASLLAYKMGITQCTIAHALEKTKYPDSDIYWKKFEEKYHFSCQ
29289943 PFRTENG-ILHKWISRFDVWPYLEKFTED--VAGEMSAELQGVPDLIIGNYSDGNLVASLLAYKMGVTQCTIAHALEKTKYPDSDIYWKKFEEKYHFSCQ
115310620 PFRTEKG-ILRKWISRFDVWPYLETFTED--AANEISAELQGRPDLIIGNYSDGNLVASLLAHKLGVTQCTIAHALEKTKYPDSDIYWRKFEEKYHFSCQ
4468153 PFRTEHG-VLRQWISRFDVWPYLEKFAED--AASEISAELRGVPDLIIGNYSDGNLVASLMAHKMGVTQGTVAHALEKXKYPNSDIYWKXYEDKYHFSCQ
345286419 PFRTELG-VLHKWISRFDVWPYLEKFAED--AAGEIAAELQGVPDLIIGNYSDGNLVASSLSHKMGVTECTIAHALEKTKYPDSDLYWKKYEEKYHFSCQ
4468151 PFRTEQG-ILKQWISRFDVWPYLEKFAED--AGKEIRAEMKAVPDLLIGNYSDGNLVASLLAYKMGVTQCTIAHALEKTKYPDSDIYWKKHEEKYHFSCQ
151176306 PFRSEKG-ILRKWISRFDVWPYLENFTED--AAGEIIGELQGRPDLIIGNYSDGNIVASLLSHKMGVTQCNIAHALEKTKYPDSDIYWKRFEDKYHFSCQ
14530225 PFRSEKG-ILRKWISRFDVWPYLENFTED--AAGEIIGELQGRPDLIIGNYSDGNIVASLLSHKMGVTQCNIAHALEKTKYPDSDIYWKRFEDKYHFSCQ
15239816 PFRTEKG-ILRKWISRFDVWPYLETFAED--ASNEISAELQGVPNLIIGNYSDGNLVASLLASKLGVIQCNIAHALEKTKYPESDIYWRNHEDKYHFSSQ
297795665 PFRTEKG-ILRKWISRFDVWPYLETFAED--ASNEISAELQGVPNLIIGNYSDGNLVASLLASKLGVMQCNIAHALEKTKYPESDIYWRNHEDKYHFSSQ
225431790 PFRTENG-ILRKWISRFDVWPYLETFAED--ASNEIAAELQGVPDLIIGNYSDGNLVASLLSYKMGITQCNIAHALEKTKYPESDIYWRKFEDKYHFSSQ
255551835 PFRTQNG-ILRKWISRFDVWPYLETFADD--ASNEIAAELQGVPDLIIGNYSDGNLVASLLSYKLGITQCNIAHALEKIKYPDSDIYWRKYEDKYHFASQ
168009716 PFRHDGK-VLNNWISRFDVYPYLETYAQE--AAREISTDLAGPPDLIIGNYSDGNLVATLMCQQLGVTQCTIAHALEKTKYPDSDIYWKKFEEKYHFSCQ
168035060 PFRNDGQ-ILKNWISRFDVYPYLETYAQE--AASEICADLSGPPDLIIGNYTDGNLVATLLCQHLGVTQCTIAHALEKTKYPDSDIYWKNFEEKYHFSCQ
168058907 PFQHEGN-ILKNWISRFDVYPYLENYAQD--AAREVLGELQGRPDLIIGNYSDGNLVATLLSHYLDVTQCIIAHALEKTKYPDSDIYWKDFEEKYHFSCQ
168029793 PFRNENG-ILHNWISRFDVYPFLENFVYD--VAQELTVELPGKPDFIIGNYTDGNLVASLLCHQLGVTQCNIAHALEKTKYPDSDIYWKKFEEKYHFSCQ
94266940 PFRYPDLSVVPHWLSRFKLWPFLDQFAVD--VEEELRQEFAGRPDLLVGNYSDGNLVATRLSTSMGVIQCNIAHALEKSKYLFSDLYWHQFEADYHFSIQ
94264333 PFRYPDLSVVPHWLSRFKLWPFLDQFAVD--VEEELRQEFVGRPDLLVGNYSDGNLVATRLSTSMGVIQCNIAHALEKSKYLFSDLYWHQFEADYHFSIQ
297569306 PFRYPDNSVVPHWLSRFRIWPYLDQFAVD--AEEEIRRELGGRPDLLVGNYSDGNLVATRLSKNMGVIQCNIAHALEKSKYLFSDLYWDEFEPEYNFSIQ
254416162 PFREFNPNYTQNWISRFEIWPYLETYAID--AEKELRAEFQGVPDFIVGNYSDGNLVAFLLSRRMDVTQCIVAHALEKSKYLFSNLYWQELEDQYHFSIQ
332712456 PFREFNPKVTQDWISRFEIWPYLETYAID--AEKELLAEFQGRPDLIVGNYSDGNLVAFLLARRLKVTQCNIAHALEKSKYLFSNLYWQDSEQQYHFSLQ
17232477 PLREFNPKMTQNWISRFEFWPYLETFAID--SERELLAEFQGRPDLIVGNYTDGNLVAFLLTRRMKVTQCNIAHALEKSKYLFSNLYWQDLEEKYHFSLQ
75908500 PLREFNPKMTQNWISRFEFWPYLETFAID--SERELLAEFHGRPDLIVGNYTDGNLVAFLLARRMKVTQCNIAHALEKSKYLFSNLYWQDLEEKYHFSLQ
186682280 PLRDFNPNMTQNWISRFEFWPYLETFAID--SERELRAEFQGTPDLIVGNYTDGNLIAFLLARRLKVTQCNVAHALEKSKYLFSNLYWQELEEKYHFSLQ
119512682 PLRDFNPNMTQNWISRFEFWPYLETFAID--SEKELLAEFQGRPDLIVGNYSDGNLVAFLLARRMDVTQCNIAHALEKSKYLFSNLYWEELEDKYHFSLQ
334117431 PFRQFNPKLTQNWISRFEIWPYLETYALD--SEKELLAEFQGKPDLIVGNYSDGNLVAFLLSRKLKTTQCNIAHALEKSKYLFSNLYWQESEDKYHFSLQ
218440696 PFRDFNPNMTQNWISRFEIWPYLETYAID--AEKELYAQFHGRPDLIIGNYSDGNLVAFLLARRLDVTQFNIAHALEKSKYLFSNLYWQDLEHLYHFSIQ
159031025 PFREFNPKLTQNWISRFEIWPYLETFALD--AEREILAEFQGRPDLIIGNYSDGNLVAFLLARKLKVTQCNIAHALEKSKYLFSNLYWQDLEEQYHFSLQ
220907171 PFRSFNPKVTENWISRFEIWPYLETFAID--AEKALLAEFAGRPDLIVGNYSDGNLVAFLLSRRLGVTQGIIAHALEKSKYLFSNLYWQDLDDKYHFSLQ
22298591 PFREFNPKVTQNWISRFEIWPYLETFAID--AERELRAEFGHVPDLIIGNYSDGNLVAFLLARRLKVTQCNIAHALEKSKYLFSNLYWQDLEDKYHFSLQ
158339122 PFRELNPKYTQNWISRFEIWPYLETYAID--AERELLAEFRGLPDLIVGNYTDGNLVAFLLSRRLGVTQCNVAHALEKSKYLFSNLYWQDLEEQYHFSMQ
284053161 PFREFNPNVTQNWISRFEIWPYLETFAID--AETEILAEFQGRPDLIVGNYSDGNLVAFLLSKRLNVIQCNVAHALEKSKYVFSDLYWQDMDDKYHFSLQ
209523126 PFREFNPNVTQNWISRFEIWPYLETFSID--AETEILAEFQGRPDLIVGNYSDGNLVAFLLSKRLNVIQCNVAHALEKSKYVFSDLYWQNMEDKYHFSLQ
17228554 PFGEFNPEITNNWISKFEIWPYLETFALD--AEKQLLAQFQGKPNLIIGNYSDGNLVAFLLARRLKVTHCNIAHSLEKPKNLFSNLYWQDSEEKYHFSVQ
75909957 PFGEFNPAITNNWISKFEIWPYLETFALD--AEKQLLAQFQGKPNLIVGNYSDGNLVAFLLARRLKVTHCNIAHSLEKPKNLFSNLYWQNSEEKYHFSVQ
186685043 PFGEFNPEITNNWISKFEIWPYLEQFAID--AEKELLTQFKGKPNLLVGNYSDGNLVASLLSRRMKVTQCNIAHSLEKPKYLFSNLYWQDLENQYHFSAQ
119510624 PFGEFNSDITNNWISKYEIWPYLETFAQD--AEKELLTQFQGRPNLIVGNYSDGNLVASLISRSLKVTQCNIAHSLEKPKHLFSNLYWQDLEDNYHFSAQ
298489784 PFRDFNPEIINNWISKFEIWPYLEKFAQD--AERELLTEFSGKPDLIVGNYSDGNLVASLLSSSLKIPQCNIAHSLEKPKHLFSNLYWQDLDEKYHFSAQ
220909283 PFRECNPNVTDNWISKFEIWPYLENFAID--AEKVLLAQLGGRPQLIVGHYSDGNLVAYLLARRFKAIHCQVAHALEKPKYLFSDLYWQDWEAQYHFSAQ
332708740 PFQEFNPKVTQNWISKFEIWPYLESFALD--SEKALLEEFQGSPDLIIGNYSDGSLVAFLLARRLNAIHGSIAHTMEKPKYLFSDLYWKDFESQYNFSIQ
Clustal Consensus *: :* : *::::*:*::*.. : : : :**::*: * * :.* *
510 520 530 540 550 560 570 580 590 600
114331077 YTADLLAMNAADFIVTSTYQEIAGTREAEGQYESYRAFSMPGLYRVINGIDLFDPKFNIVSPGADAEVYFPYTDQ-SRRLHSLIPEIESMLFDNTANFPA
30249199 YTADLLAMNSADFIVTSTYQEIAGTREAEGQYESYQAFSMPDLYRVIHGIDLFDPKFNIVSPGANADIYFPYSDP-NRRLHSLIPEIESLIFDDATNLPA

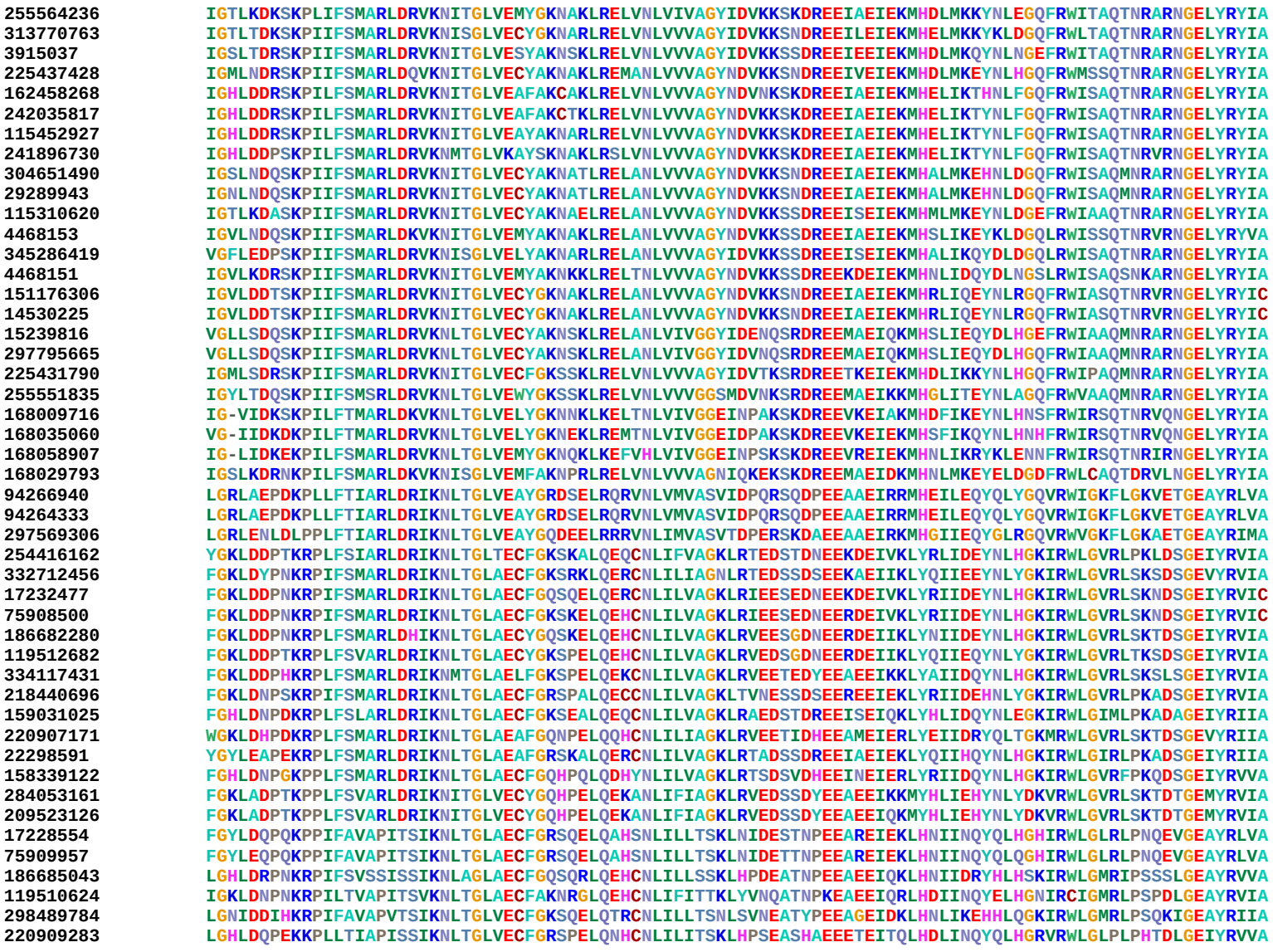
```
82703384        YTADLIAMNSADFIITSTFQEIAGTEQTVGQYETYQNYTMPGLYRVVNGIDLFDPKFNIVSPGADAEVYFSYLDH-ERRLDALIPDIERLLYGDDPGVPC
198283392       FTADLIAMNSSDIIVTSTYQEIAGNDREVGQYEGYQNYSLPGLYRVENGIDVFDTKFNIVSPGADAHYYFPYSAS-EARLRYLHDDIDALLFGEEPAADR
344199811       FTADLIAMNSADIIVTSTYQEIAGNDHEVGQYEGHQNYSLPGLYRVENGIDVFDTKFNIVSPGADAHYYFPYSAS-EERLRYLHDDIDALLFGEEPAADR
255021595       FTADLIAMNAADIIVTSTYQEIAGNDREIGQYEGHQDYTLPGLYRVENGIDVFDSKFNIVSPGADPRFYFSYART-EERPSFLEPEIESLLFGREPGADR
77166514        FTGDLIAMNSADFIVTSTYQEIAGNKNSVGQYESYSAYTLPGLYQVIHGIDVFDPKFNIVSPGADGEVYFPYTDT-KRRLSGLRQEIEALIWGDERP-DA
300115586       FTGDLIAMNSADFIITSTYQEIAGNKNSVGQYESYSAYTLPGLYQVIHGIDVFDPKFNIVSPGADGEVYFPYTDT-KRRLSGLRQEIEALVWGDDRS-DT
292493898       FTADFIAMNSADFIITSTYQEIAGDRSSVGQYESYGAYILPGLYQVVQGIDVFDPKFNIVSPGADAEVYFPYRER-KRRLRGLRREIEELIWGNGRP-DA
350552412       FTADLIAMNAADFIITSTYQEIAGTDHSVGQYESYSAFSMPGLYRVVKGIDVFDPKFNIVSPGADAEVYFSYKDS-ERRLHGLHDELQTLIFGTPSE-DM
220933364       FTADLIAMNAADFIITSTYQEIAGTGEDIGQYESYMSFSLPDLYRVVRGIDVFDPKFNIVSPGADDRVYFPYTEE-ERRITGLHEEIEALLFGGHRD-DA
350574485       FTADLIAMNTADFIITSTYQEIAGTDESLGQYESYMNFTMPGLYRVVAGVDVYDPKFNIVSPGADEEIYFPFTET-ERRLAHLHGEIEQLIFGEPVPGQS
95929190        FTADLVSMNAADFIITSTYQEIAGTEESLGQYESYSSFTMPALYRVINGINIYDPKFNIVSPGADDRVYFPYYDE-ENRLTELHDELHELIYGDHME-GS
298528445       FTADLIAMNTADFIITSTYQEIAGTEESVGQYETYNAFTMPDLYRVVSGIDVFDPKFNVVSPGADENVYFPYYEK-DRRLTELHDELSDYIYGPPGD-WA
297602308       FTADMISMNTSDFIITSTYQEIAGSKEKPGQYEHHYAFTMPGLCRYATGINVFDPKFNIAAPGADQSIYFPFTQK-QKRLTDLHPQIDELLYSKDDTDEH
115457664       FTADMISMNTSDFIITSTYQEIAGSKEKPGQYEHHYAFTMPGLCRYATGINVFDPKFNIAAPGADQSIYFPFTQK-QKRLTDLHPQIDELLYSKDDTDEH
326504012       FTADMFAMNTTDFIITSTYQEIAGSKEKPGQYEHHYAFTMPGLCRFATGINVFDPKFNIAAPGADQSVYFPFTQK-QKRLTNLHPQIEELLYSKEDTDEH
115450038       FTADMIAMNTSDFIIASTYQEIAGSKEKPGQYESHYAFTMPGLCRYATGINVFDPKFNIAAPGADQSVYFPFTQK-QKRLTDLHPQIEELLYSKEDNNEH
326531526       FTADMIAMNTSDFIIASTYQEIAGSKDKPGQYESHYAFTMPGLCRYATGVNVFDPKFNIAAPGADQTVYFPFTQK-QARLTDLHPQIEELLYSKEDNDEH
242063616       FTADMIAMNTSDFIIASTYQEIAGSKDKPGQYESHYAFTMPGLCRFATGINVFDPKFNIAAPGADQSVYFPFTLK-HKRLTDLHPQIEALVYGKEENDEH
42568160        FTADLISMNSADFIIASTYQEIAGSKERAGQYESHMSFTVPGLYRVVSGINVFDPRFNIAAPGADDSIYFPFTAQ-DRRFTKFYTSIDELLYSQSENDEH
297805240       FTADLISMNSADFIIASTYQEIAGSKERVGQYESHMSFTLPGLYRVVSGINVFDPRFNIAAPGADDTIYFPFTAQ-DRRFTKFYPSIEELLFSQNENDEH
313770765       FMADIVAMNATDFIIASTYQEIAGSKDRTGQYESHAAFTLPGLCRVVSGVNVFDPKFNIAAPGADQSVYFPHTEK-QSRFTQFNPDIEELLYSKVVNDEH
313770767       FMADTIAMNATDFIIASTYQEIAGSKDRPGQYESHASFTLPGLCRVVSGIDVFDPKFNIAAPGADQSVYFPYTEK-QSRFTKFHPAIEELLYSKVVNDEH
255584097       FIADTISMNAADFIIASTYQEIAGSKERPGQYESHSAFTLPGLCRVVSGINVFDPKFNVAAPGADQSVYFPNTEK-QKRFSQFHSAIEELLYSKEENEEH
225466221       FTADTISMNAADFIITSTYQEIAGSKDRPGQYESHTSFTLPGLCRVVSGINLFDPKFNIAAPGADQSVYFPYMER-HKRLTSFQPAIEELLYSKQDNNEH
313770769       FTADMIAMNSADFIITSTYQEIAGSNVRPGQYESHTAFTMPGLCRVVSGINVFDPKFNIASPGADQSVYFPYTEK-QKRLTSFHPAIEELLYSNEDNHEH
313770771       FTADMIAMNTADFIITSTYQEIAGSKNRPGQYESHVAFTMPGLCRVVSGINVFDPKFNIASPGADQTVYFPYTEK-QKRLTSFHPAIEELLYNNEDNNEH
255570671       FTADMIAMNAADFIITSTYQEIAGSKDRPGQYESHKAFTMPGLCRVVSGVNVFDPKFNIAAPGADQSVYFPYTEK-RRRLTSFYPAIEELIYSKEGNDEH
15219457        FTADLIAMNVTDFIITSTYQEIAGSKDRPGQYESHTAFTMPGLCRVVSGIDVFDPKFNIAAPGADQSVYFPYTEK-DKRFTKFHPSIQELLYNEKDNAEH
297842089       FTGDLIAMNVTDFIITSTYQEIAGSKDRPGQYESHTAFTMPGLCRVVSGIDVFDPKFNIAAPGADQSVYFPYTEK-EKRFTKFHPSIQELLYNEKDNAEH
115466896       FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHIAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSVYFPYTEA-DKRLTAFHPEIEELLYSEVENDEH
17980243        FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHIAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSVYFPYTET-DKRLTAFHPEIEELIYSDVENSEH
162460741       FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHIAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSVYYPYTET-DKRLTAFHPEIEELIYSDVENSEH
239984690       FTADLIAMNHTDFIITSTFQEIAGSKDSVGQYESHIAFTLPDLYRVVHGIDVFDPKFNIVSPGADMTVYFPYTET-DKRLTAFHSEIEELLYSDVENDEH
162460681       FTTDLIAMNHADFIITSTFQEIAGNKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADLSIYFPYTES-HKRLTSLHPEIEELLYSQTENTEH
242035533       FTTDLIAMNHADFIITSTFQEIAGNKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADLSIYFPYTES-HKRLTSLHPEIEELLYSQTENTEH
62865493        FTTDLIAMNHADFIITSTFQEIAGNKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADLSIYFPYTES-HKRLTSLHPEIEELLYSDVDNHEH
17980241        FTTDLIAMNHADFIITSTFQEIAGNKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYSES-HKRLTSLHPEIEELLYSDVDNNEH
115453437       FTTDLIAMNHADFIITSTFQEIAGNKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYSES-RKRLTSLHPEIEELLYSEVDNNEH
115473359       FTADLIAMNHADFIITSTFQEIAGNKETVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPFTES-QKRLTSLHLEIEELLFSDVENTEH
3980298         FTADLIAMNHADFIITSTFQEIAGKKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYTEQ-QKRLTSLHTEIEELLFSDVENAEH
15242073        FTADIFAMNHTDFIITSTFQEIAGSKETVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYTEE-KRRLTKFHSEIEELLYSDVENKEH
297812265       FTADIFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYTEE-KRRLTKFHSEIEELLYSDVENEEH
22331535        FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHRSFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFAYTEE-KRRLTAFHLEIEELLYSDVENEEH
297818772       FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMNIYFAYTEE-KRRLTAFHQEIEELLYSDVENEKH
313770759       FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADESIYFPYTDE-KRRLTSFHPEIDELLYSPVENEEH
313770761       FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADESIYFPYTEK-KLRLTSFHEEIEELLYSSVENDEH
255550319       FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADESIYYPYTDT-KRRLTSFHPEIEELLYSPVENEEH
```

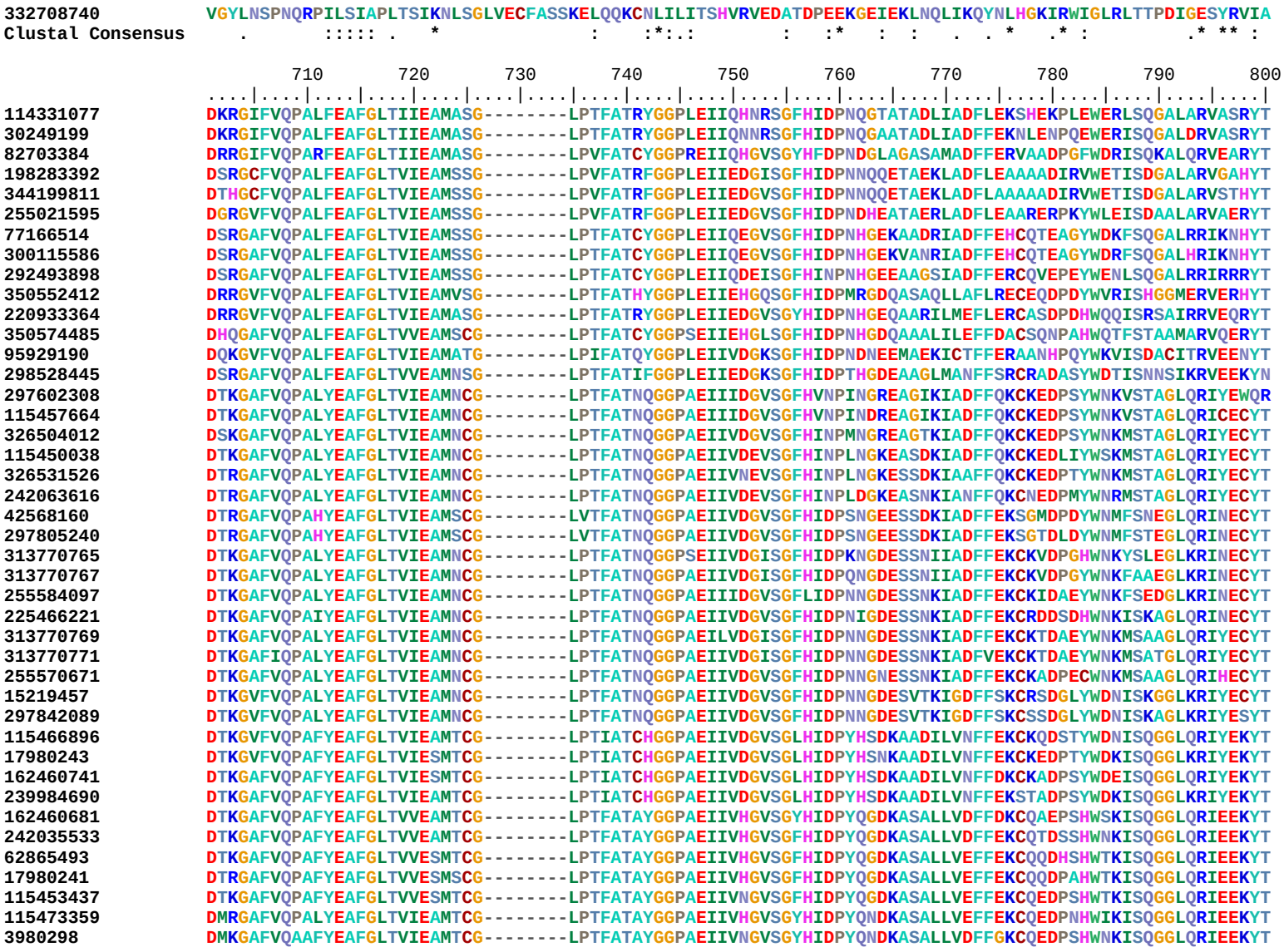
```
324984223       FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMEIYFPYTEE-KRRLKHFHTEIEDLLYSKVENEEH
258489633       FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMEIYFPYTEE-KRRLKHFHTEIEDLLYSKVENEEH
324984221       FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMEIYFPYTEE-KRRLKHFHPEIEDLLYSKVENEEH
324984229       FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMEIYFPYTEE-KRRLKHFHPEIEDLLYSKVENEEH
6683114         FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYTEE-KRRLKSFHPEIEELLYSDVENKEH
4584690         FTADLFAMNHTDFIITSTFQEIAGSKDKVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADQTIYFPYTET-SRRLTSFYPEIEELLYSSVENEEH
3915046         FTADLFAMNHTDFIITSTFQEIAGSKDKVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADQTIYFPYTET-SRRLTSFYPEIEELLYSSVENEEH
3766299         FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADQTIYFPYTET-SRRLTSFYPEIEELLYSTVENEEH
3377764         FTADLFAMNHTDFIITSTFQEIAGSKDTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADQTIYFPYTET-SRRLTSFYPEIEKLLYSTGGNEEH
225444613       FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHTGFTMPGLYRVVHGIDVFDPKFNIVSPGADMTIYFSYTEE-KMRLKALHPEIEELLFSPVENKEH
3660531         FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHMAFTMPGLYRVVHGINVFDPKFNIVSPGADINLYFPYSES-EKRLTAFHPEIDELLYSDVENDDH
31455440        FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHMAFTMPGLYRVVHGINVFDPKFNIVSPGADINLYFSYSET-EKRLTAFHPEIDELLYSDVENDEH
304651488       FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADVNLYFPYSEK-EKRLTTFHPEIEDLLFSDVENEEH
28629438        FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADVNLYFPYSEK-EKRLTTFHPEIEDLLFSDVENEEH
115310618       FTADLIAMNHTDFIITSTFQEIAGSKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADTNLYFPHTEK-EKRLTSFHPEIEELLFSDVENEEH
345286417       FTADLYAMNHTDFIITSTFQEIAGSKDTVGQYESHMAFTMPGLYRVVHGIDVFDPKFNIVSPGADMNLYFPYTEK-EKRLTALHPEIEELLYSNVENDEH
38425095        FTADLIAMNHTDFIITSTFQEIAGNKDTVGQYESHMAFTLPGLYRVVHGIDVFDPKFNIVSPGADLSIYFNYTEE-KKRLTALHPEIEELLFSETQNEEH
15235300        FTADLIAMNNADFIITSTYQEIAGTKNTVGQYESHGAFTLPGLYRVVHGIDVFDPKFNIVSPGADMTIYFPYSEE-TRRLTALHGSIEEMLYSPDQTDEH
297814081       FTADLIAMNNADFIITSTYQEIAGTKNTVGQYESHGAFTLPGLYRVVHGIDVFDPKFNIVSPGVDMAIYFPFSEE-TKRLTALHSSIEEMLYSPEQTDEH
6682841         FTADLIAMNNADFIITSTYQEIAGTKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMDIYFPYSEK-QKRLTALHGSIEQLLFDPEQNDEH
255564236       FTADILAMNNADFIITSTYQEIAGSKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYSEK-QKRLTALHGSIEKMLYDPEQTDEW
313770763       FTADVLAMNNADFIITSTYQEIAGTKTTVGQYESHTAFTLPGLYRVVHGINVFDTKFNIVSPGADMDIYFPYSDK-QKRLTTLHGSIEKMLYDSEQTDDW
3915037         FTADLIAMNNADFIITSTYQEIAGTKNTIGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMTIYFPYSDK-EKRLTALHSSIEKLLYGTEQTDEY
225437428       FTADLIAMNNADFIITSTYQEIAGSKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMCIYFPYSDV-EKRLTALHGSIEKLLYDPEQNEEH
162458268       FTADIIAMNNADFIITSTYQEIAGSKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPHTEK-AKRLTSLHGSIENLIYDPEQNDEH
242035817       FTADIIAMNNADFIITSTYQEIAGSKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPHTEK-AKRLTSLHGSIENLLYDPEQNDQH
115452927       FTADIIAMNNADFIITSTYQEIAGSKNTVGQYESHTAFTLPGLYRIVHGIDVFDPKFNIVSPGADMSIYFPYTEK-AKRLTSLHGSLENLISDPEQNDEH
241896730       FTADIIAMNNADFIITSTYQEIAGSKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPFTEK-AKRLTALHGSIESLIYDPEQNDEH
304651490       FTADLLSMNHSDFIITSTYQEIAGTKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMTIYFPYFDK-EKRLTSLHPSIEKLLFDPEQNEVH
29289943        FTADLLSMNHSDFIITSTYQEIAGTKNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMTIYFPYSDK-EKRLTSLHPSIEKLLFDPEQNEVH
115310620       FTADLLAMNHSDFIITSTYQEIAGTNNTVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMAIYFPYSDT-EKRLTSFHGSIENLLFDPEQNDEH
4468153         FTADLLAMNNSDFIITSTYQEIAGTKNSVGQYESHAGFTLPGLYRVVHGIDVFDPKFNIVSPGADDGIYFSYSEK-ERRLTSYHDCLEKLLFDPQQTEEH
345286419       FTADLLAMNHSDFIITSTYQEIAGTKSTVGQYESHASFTLPGLYRVVHGIDVFDPKFNIVSPGADECIYFPYSEK-DKRLTALHESLEKLIFDPQQTDEH
4468151         FTADLLAMQHSDFIITSTYQEIAGTRNVVGQYESHVAFTMPGLYRVVHGIDVFDPKFNIVSPGADESIYFPYTDK-EKRLTNLQASIEKLLFDPEQNEEH
151176306       FSADLMAMNHADFIITSTYQEIAGTKNTVGQYESHKAFTFPGLYRVVHGIDVFDPKFNIVSPGADMAIYFPFSEK-EHRLTSLHSFIEQLLFKPEQNEEH
14530225        FSADLMAMNHADFIITSTYQEIAGTKNTVGQYESHKAFTFPGLYRVVHGIDVFDPKFNIVSPGADMAIYFPFSEKDVTCLTSLHRLIEQLLFKPEQNEEH
15239816        FTADLIAMNNADFIITSTYQEIAGSKNNVGQYESHTAFTMPGLYRVVHGIDVFDPKFNIVSPGADMTIYFPYSDK-ERRLTALHESIEELLFSAEQNDEH
297795665       FTADLIAMNNADFIITSTYQEIAGSKNNVGQYESHTAFTMPGLYRVVHGIDVFDPKFNIVSPGADMTIYFPYSDK-ERRLTALHESIEELLFSAEQNDEH
225431790       FTADLIAMNNADFIITSTYQEIAGSKNHVGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFSYSEK-ERRLTALHDSIESLLYDSEQNDDH
255551835       FTADIIAMNNADFIITSTYQEIAGNKNNIGQYEGYTAFTLPGLYRVVHGINVFDPKFNIVSPGADSCIYFPYSDR-ERRLTALHGAIEELLYDPEQNEEH
168009716       FTADLIAMNHADFIITSTYQEIAGSAKTVGQYESHQAFTMPGLYRVVNGVNVFDPKFNIVSPGADMDVYFPYTDK-ERRLTKLHPTIEDLLFGTEQSDEH
168035060       FTADLIAMNHADFIITSTYQEIAGSAKTVGQYESHQAFTMPSLYRVVNGIDVFDPKFNIVSPGADMTVYYPFTDK-QHRLTKLHPAIEKLLFSSDQTDEH
168058907       FTADLIAMNSADFIITSTYQEIAGSADTVGQYESHQAFTMPGLYRVVNGIDVFDPKFNIVSPGADMNIYYPFADK-ERRLTSLQESIEELLYSPEQTDEH
168029793       FTADLIAMNQADFIITSTYQEIAGSEDTVGQYESHVAFSLPGLYRVVNGIDVFDPKFNIVSPGADTIVYFSFTEK-DRRLTDLHDKIEKLLYDPEQTAEH
94266940        FMVDLIAMNQANFIVTSTAQEITGTENSIGQYESYQFFTMPGLLNITSGIDLFHPRFNVIPPGVNQEVYFPWNRK-RSRPTKLRRRVSELLFSGEDD-DC
94264333        FMVDLIAMNQANFIVTSTAQEITGTENSIGQYESYQFFTMPGLLNITSGIDLFHPRFNVIPPGVNQEVYFPWNRK-RSRPTKLRRRVSELLFSGDDD-DC
297569306       FMADLLAMNQANFIITSTAQEITGTDNSIGQYESYQFFTMPGLVNVISGINLFHPRFNVIPPGVNQEVYFPYNRK-RGRKVKMRREVTRLLFEQEDA-DC
```

```
254416162          FTADLIAMNAANFIISSTYQEIVGKPDSVGQYESYQNFTMPDLYHVVNGIELFSPKFNVVPPGVNEQVYFPYTRS-EDRVPRDCERLEELLFTLDDPSQV
332712456          FTADLIAMNAANFIISSTYQEIVGTADSVGQYESYQNFTMPDLYHVINGIELFSPKFNVVPPGVNETVFFPYTRT-QGRVASDIKRLDEFLFTLDDEAQV
17232477           FTADLIAMNAANFVISSTYQEIVGTPDSIGQYESYKCFTMPELYHVVNGIELFSPKFNVVPPGVNENSYFPYTQT-QNRIESDRDRLEEMLFTLEDSSQI
75908500           FTADLIAMNAANFVISSTYQEIVGTSDSVGQYESYKCFTMPELYHVVNGIELFSPKFNVVPPGVNENSYFPYTHT-QDRIESDRDRLEEMLFTLEDSSQI
186682280          FTADLIAMNAANFVVSSTYQEIVGTPDSVGQYESYKCFTMPELYHVTNGIELFSPKFNVVPPGVNENNYFPYTRT-KDRVESDRQRLAETLFTLEDPTQI
119512682          FTADLIAMNAANFVVSSTYQEIVGTPDSVGQYESYKCFTMPELYHVVNGIELFSPKFNVVPPGVNESYYFPYTRK-EDRVEADSDRLADILFTLEDPHQI
334117431          FTADIIAMNAANCIVSSTYQEIVGQPDSVGQYESYHCFTMPDLYHVVNGIELFSPKFNVVPPGVNESVYFPYTRI-EDRVQGDRDRLNELLFTLEDPEQV
218440696          FTADLIAMNAANCIISSTYQEIVGRTDSVGQYESYESFTMPNLYHVVNGIELFSPKFNVVPPGVNENVYFPYTRT-DERVPNKREHLEDLLFTLEDPSQV
159031025          FTADLISMNAANFIISSTYQEIVGTSDSVGQYESYQCFTMPELYHVVSGIELFSPKFNVVPPGVNESYFFPYTNR-EERLLGEGERLEELLFTLEAPRRV
220907171          FTADLIVMNAAHFIISSTYQEIVGTPDSVGQYESYSSFTMPELYHVISGIELFSPKFNLVPPGVNENYFFPYTKV-EERLVTERHRLEELLFTLDDPAQV
22298591           FTADLIAMNAANFIISSTYQEIVGTPDSIGQYESYQSFTMPDLYHVVNGIELFSPKFNVVPPGVNEQVYFPYYHY-TERLEGDRQRLEELLFTLEDPQQI
158339122          FTADLIAMNAANFIVTSTYQEIAGRPDTIGQYESYRSFTMPDLYHVVYGAELFSPKFNVVPPGVNESVYFPFTRH-QERTPGDIDRLEELLFTLEDPEHV
284053161          FTADLIAMNAANFIISSTYQEIIGTTDSVGQYESYKSYTMPGLYHVVNGIELFSPKFNVVPPGVNETIFFPYTRT-EERISSDRQRLEELIFYLDDPNQV
209523126          FTADLIAMNAANFIISSTYQEIIGTTDSVGQYESYKSYTMPGLYHVVNGIELFSPKFNVVPPGVNETIFFPYTRT-EERISSDRQRLENLIFHLDDPSQV
17228554           FTADLITMNAADFIITSSYQEIFGTPESIGQYESYKFFTMPHLYHVVDGIDLFNPKFNMVPPGVNEQVFFPYSQT-ADRDPNVSKHVHDLLFHRQDS-QI
75909957           FTADLITMNAADFIITSSYQEIFGTPESVGQYESYKFFTMPHLYHVVDGVDLFSPKFNMVPPGVNEQVFFPYSQT-ADRDPNLSQSVHDLLFHRQDS-QI
186685043          FTADLISMNAADFIITSSYQEIVGTPDTIGQYESYKCFTMPQLYHVVDGIDLFSPKFNLVPPGVNESIFFPYSQK-ENRDSNLCTEIHNLLFSREDP-QI
119510624          FTADLISMNAADFIITSSYQEIVGTPDTIGQYESYKCFTMPELYHVINGIDLFSPKFNMVPPGVSENIFFPFSQK-DRRNPKLTSQVHDLLFEREHP-QI
298489784          FTADLISMNAADFIVASSYQEIVGTPDTMGQYESYKCFTMPQLYHVVDGIDLFNPKFNMIPPGVSETLFFPYSQT-EDRNPQESQKIKDLLFQQQDD-HI
220909283          FTADLITMNAANFIITSSYQEIVGTPESVGQYESYKCFTMPQLYHVVNGIELCSPKFNRVPPGVNENIFFPYTHK-DSRDPCLCQRIHHLLFHHTDD-HI
332708740          FTADLIAMNSADFILTSTYEELVGTPESVGYYESYKCFSMPELYHVVNGIELFSPKFNVVPPGVNENIFFPYTQT-SDRIAHDSERVKDLLLSKEDP-EI
Clustal Consensus  :  *   *: :. :::*: :*: *     * ** :  : .* * .   * ::  .:**  .**..   ::               :   :

                                610       620       630       640       650       660       670       680       690       700
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
114331077          RGILQDSDKPLIFTMARLDRIKNITGLVESYGASQRLRSLANLVIVGGKIDPQHSSDHEEQEQIHQMHHLMDEYKLDPQVRWLGMRLDKNLAGELYRYIA
30249199           RGYLQDPDKPLIFTMARLDRIKNITGLVELYAASPRLRSLANLVIVGGKIDPQHSSDHEEQEQIHRMHQLMDEHELDQQVRWLGMRLDKNLAGELYRYIA
82703384           RGYFADPAKPLIFTMARLDTVKNLTGLAAWFGQCEALSTAANLLVIGGHIDPAASCDGEERAEIEHMHALMNEYKLEGRMRWLGTRLEKNLAGELYRHVA
198283392          RGVLKERDKPIIFSMARMDHIKNLSGLAEIFGASERLRKLANLVIIGGHVDLQNSQDEEEGAQIQRMHDIMDAHQLDGQMRWIGTLLEKNVAGELYRVIG
344199811          RGVLKDRDKPIIFSMARMDHIKNLSGLAALFGASERLRTLANLVIIGGHVDPQDSQDEEERAQIHRMHGIMDAHQLDGQMRWIGTLLDKNVAGELYRVIG
255021595          RGVLEDRQKPLLLSMARMDRIKNLSGLAELYGRSSRLRGLANLVIIGGHVDVGNSRDAEEREEIRRMHEIMDHYQLDGQLRWVGALLDKTVAGELYRVVA
77166514           RGKLQDHTKPLLFTIARLDRIKNITGLVEWYGRCERLRKLANLVVVGGYIDKSQSADSEEQVQIARMHQLIEEYKLDSQVRWLGVMLQKNLAGELYRFIA
300115586          RGKLQDRSKPLLFTIARLDRIKNITGLVEWYGRCERLRQLVNLVVVGGYIDKSQSADSEEQAQIARMHQLMEEYGLDGQVRWLGVMLQKNLAGELYRFIA
292493898          RGRLQDKGKPLLFTMARLDRIKNITGLVEWYGRCERLRKQVNLVVVAGYVDEAQSADSEEQAQIARMHQLMEEYELDNQVRWLGTLLQKNLAGELYRFVA
350552412          RGTLKHPERPLIFTMARLDRIKNIAGLVQWYAENEALREQANLLIVAGYTDAGKSADREEQEQIGYLHHLFTTHGLDEQVRWLGVRLDKVFAGELYRFIA
220933364          RGVLAAPERPVIFTMARLDRIKNIAGLVSWYAQNAELRARANLVVVAGTVDPSRSDDQEEQAQIARMHQLFDAHDLNDCVRWLGVRLDKTLSGELYRCIA
350574485          RGQLQDRDKPLLFSMARLDRIKNIGGLVDWYARAPELRNRVNLVVVAGHVDGNASGDDEEREQIDYIHYLMNTHGLDGQVRWLGVHLDKFLAGEFYRCIA
95929190           RGLLDDKDKPLIFTMARLDKVKNITGLVECYAKSERLREQANLLVVAGSIHVDHSSDAEERYQIETMHRLFDEYQLDGQVRWLGKHLQKNKAGELYRYIA
298528445          KGELQDRTKPILFTMARLDRIKNLTSLVRWYGENPELRQEANLVLVAGSLDVRDSQDEEEKACIQEMHRLFEEFDLHEQVRWLGTRLDKNMSGELYRFIA
297602308          IGYLADRNKPIIFSMARLDKVKNITGLVEWYGQNKKLRDLVNLVVVAGLLDASQSKDREEIEEINKMHNLMDRYQLKGQIRWIKAQTDRVRNGELYRCIA
115457664          IGYLADRNKPIIFSMARLDKVK----------NKKLRDLVNLVVVAGLLDASQSKDREEIEEINKMHNLMDRYQLKGQIRWIKAQTDRVRNGELYRCIA
326504012          IGYLADRSKPIIFSMARLDKVKNITGLVEWYGQNKKVRDLVNLVVVAGLLNAAQSKDREEIDEINKMHNLIDKYQLKGQIRWIKAQTDRVRNGELYRYIA
115450038          IGHLADRSKPIIFSMARLDKIKNITGLVEWYGQNKRLRDLVNLVIVGGLLDPSQSKDREEIEEINKMHSLINKYQLVGQIRWIKGQTDRVRNGELYRCIA
326531526          LGYLGDRSKPIIFSMARLDKVKNITGLVEWYGENKKLRDLVNLVIVGGLLEPSQSNDREEIEEINKMHSLMDKYQLKGQIRWIKAQTERVRNGELYRCIA
242063616          IGYLENRRKPVIFSMARLDKVKNITGLVEWYGQDKRLRVLVNLVVVGGLLDPTQSKDREEIEEINKMHSLINKYQLKGQIRWIKAQTDRVRNGELYRCIA
42568160           IGYLVDKKKPIIFSMARLDVVKNLTGLTEWYAKNKRLRDLVNLVIVGGFFDASKSKDREEISEIKKMHSLIEKYQLKGQFRWITAQTDRTRNGELYRSIA
297805240          IGYLVDKKKPIIFSMARLDVVKNLTGLTEWYAKNKRLRDLVNLVIVGGFFDPSKSKDREEISEIKKMHSLIEKYQLKGQFRWIAAQTDRTRNGELYRCIA
```

```
313770765        IGYLEDKKKPIIFSMARLDTVKNLTGLTEWYGKNKRLRGLVNLVIVGGFFDPNKSKDREEMAEIKKMHELIEKYQLKGQIRWIAAQTDRKRNGELYRCIA
313770767        IGYLEDKKKPIIFSMARLDTVKNLTGLTEWYGKNKRLRGLVNLVIVGGFFDPNKSKDREEMAEITKMHGLIKKYRLNGQFRWIAAQTDRNRNGELYRCIA
255584097        IGYLADKKKPIIFSMARFDTVKNLTGLTEWYGKNKRLRNLVNLVIVGAFFDPSKSKDREEMAEIKKMHALIDKYQLKGQIRWIAAQTDRQRNGELYRCIA
225466221        IGFLADRKKPIIFSMARLDIVKNITGLTEWFGNNKRLRSLVNLVIVAGFFDPSKSKDREEMAEIKKMHTLIEKYQLKGQIRWIAAQNDRRRNGELYRCIA
313770769        IGYLADRKKPIIFSMARLDTVKNITGLTEWFGKNTKLRNLVNLVVVAGFFDPSKSNDREEIAEIKKMHALIEKYQLKGQFRWIAAQTDRYRNGELYRCIA
313770771        IGYLADKKKPIIFSMARLDTVKNITGLTEWYGKNAKLRNLVNLVVVAGFFDPSKSNDREEIAEIKKMHSLIEKYQLKGQFRWIAAQSDRYRNGELYRCIA
255570671        IGYLADRKKPIIFSMARLDTVKNITGLTEWYGKNKRLRNLVNLVVVAGFFDPSKSKDREEIAEINKMHALIEKYQLKGQIRWIAAQTDRYRNGELYRCIA
15219457         MGYLADREKPIIFSMARLDTVKNITGLVEWYGKDKRLREMANLVVVAGFFDMSKSNDREEKAEIKKMHDLIEKYKLKGKFRWIAAQTDRYRNSELYRCIA
297842089        MGYLAEREKPIIFSMARLDTVKNITGLVEWYGKDKRLREMANLVVVAGFFDMSKSNDREEKAEIKKMHDLIEKYKLKGKFRWIAAQTDRYRNSELYRCIA
115466896        KFVLKDKNKPIIFSMARLDRVKNMTGLVEMYGKNAHLRDLANLVIVCGDHG-NQSKDREEQAEFKKMYGLIDQYKLKGHIRWISAQMNRVRNGELYRYIC
17980243         KFVLKDKNKPIIFSMARLDRVKNMTGLVEMYGKNAHLRDLANLVIVAGDHG-KESKDREEQAEFKRMYSLIDEYKLKGHIRWISAQMNRVRNGELYRYIC
162460741        KFVLKDKKKPIIFSMARLDRVKNMTGLVEMYGKNARLRELANLVIVAGDHG-KESKDREEQAEFKKMYSLIDEYKLKGHIRWISAQMNRVRNGELYRYIC
239984690        KFVLKDRNKPIIFSMARLDRVKNMTGLVEMYGKNAHLKDLANLVIVAGDHG-KESKDREEQAEFKRMYSLIEEYKLKGHIRWISAQMNRVRNGELYRYIC
162460681        KFVLNDRNKPIIFSMARLDRVKNLTGLVELYGRNKRLQELVNLVVVCGDHG-NPSKDKEEQAEFKKMFDLIEQYNLNGHIRWISAQMNRVRNGELYRYIC
242035533        KFVLNDRNKPIIFSMARLDRVKNLTGLVELYGRNKRLQELVNLVVVCGDHG-NPSKDKEEQAEFKKMFDLIEQYNLNGHIRWISAQMNRVRNGELYRYIC
62865493         KFVLKDRNKPIIFSMARLDRVKNLTGLVELYGRNPRLQELVNLVVVCGDHG-NPSKDKEEQAEFQKMFDLIEQYNLNGHIRWISAQMNRVRNGELYRYIC
17980241         KFVLKDRNKPIIFSMARLDRVKNLIGLVELYGRNPRLQELVNLVVVCGDHG-NPSKDKEEQAEFKKMFDHIEQYNLNGHIRWISAQMNRVRNGELYRYIC
115453437        KFMLKDRNKPIIFSMARLDRVKNLTGLVELYGRNPRLQELVNLVVVCGDHG-NPSKDKEEQAEFKKMFDLIEQYNLNGHIRWISAQMNRVRNGELYRYIC
115473359        KFVLKDKKKPIIFSMARLDHVKNLTGLVELYGRNPRLQELVNLVVVCGDHG-KESKDKEEQAEFKKMFNLIEQYNLNGHIRWISAQMNRVRNGELYRYIC
3980298          KFVLKDKKKPIIFSMARLDRVKNMTGLVEMYGRNPRLQELVNLVVVCGDHG-KVSKDKEEQVEFKKMFDLIEKYNLSGHIRWISAQMNRVRNGELYRYIC
15242073         LCVLKDKKKPILFTMARLDRVKNLSGLVEWYGKNTRLRELANLVVVGGDRR-KESKDNEEKAEMKKMYDLIEEYKLNGQFRWISSQMDRVRNGELYRYIC
297812265        LCVLKDKKKPILFTMARLDRVKNLSGLVEWYGKNTRLRELANLVVVGGDRR-KESKDNEEKAEMKKMYDLIEEYKLNGQFRWISSQMNRVRNGELYRYIC
22331535         LCVLKDKKKPIIFTMARLDRVKNLSGLVEWYGKNTRLRELVNLVVVGGDRR-KESQDNEEKAEMKKMYELIEEYKLNGQFRWISSQMNRVRNGELYRYIC
297818772        LCVLKDKKKPILFTMARLDRVKNLSGLVEWYGKNTRLRELVNLVVVGGDRS-KESQDNEEKAEMKKMYDLIEEYKLNGQFRWISSQMNRVRNGELYRYIC
313770759        LCVLKDRNKPILFTMARLDRVKNLSGLVEWYGKNTKLRELVNLVVVGGDRR-KESKDLEEQAEMKKMYSHIEKYNLNGQFRWISSQMNRVRNGELYRYIC
313770761        LCVLKDRNKPILFTMARLDRVKNLTGLVEWYGKNTKLRELANLVVVGGDRR-KESKDIEEQAEMKKMYSHIEKYKLNGQFRWISSQMNRVRNGELYRYIC
255550319        LCVLKDRSKPIIFTMARMDRVKNLTGLVEWYGKNAKLRELANLVVVGGDRR-KESKDLEEQAEMKKMHGLIEKYNLNGQFRWISSQMNRVRNGELYRCIC
324984223        LCVLNDRNKPILFTMARLDRVKNLTGLVEWYGKNAKLRELANLVVVGGDRR-KESKDLEEKAEMKKMFELIEKYNLNGQFRWISSQMNRIRNGELYRYIC
258489633        LCVLNDRNKPILFTMARLDRVKNLTGLVEWYGKNAKLRELANLVVVGGDRR-KESKDLEEKAEMKKMFELIEKYNLNGQFRWISSQMNRIRNGELYRYIC
324984221        LCVLNDRNKPILFTMARLDRAKNLTGLVEWYGKNAKLRELANLVVVGGDRR-KESKDLEEKAEMKKMFELIDKYNLNGQFRWISSQMNRIRNGELYRYIC
324984229        LCVLNDRNKPILFTMARLDRVKNLTGLVEWYGKNAKLRELANLVVVGGDRR-KESKDLEEKAEMKKMFELIDKYNLNGQFRWISSQMNRIRNGELYRYIC
6683114          LCVLKDRNKPILFTMARLDRVKNLTGLVEWYGKNAKLRELVNLVVVGGDRR-KESKDLEEQAEMKKMYSLIDQYKLNGQFRWISSQMNRVRNGELYRYIC
4584690          ICVLKDRNKPIIFTMARLDRVKNITGLVEWYGKNAKLRELVNLVVVAGDRR-KESKDLEEIAEMKKMYGLIETYKLNGQFRWISSQMNRVRNGELYRVIC
3915046          ICVLKDRNKPIIFTMARLDRVKNITGLVEWYGKNAKLRELVNLVVVAGDRR-KESKDLEEIAEMKKMYGLIETYKLNGQFRWISSQMNRVRNGELYRVIC
3766299          ICVLKDRSKPIIFTMARLDRVKNITGLVEWYGKNAKLRELVNLVVVAGDRR-KESKDLEEKAEMKKMYEHIETYKLNGQFRWISSQMNRVRNGELYRVIC
3377764          ICVLKDRNKPIIFTMARLDRVKNITGLVEWYGKNAKLRELVNLVVVAGDRR-KESKDLEEKAEMKKMYEHIETYKLNGQFRWISSQMNRVRNGELYRVIC
225444613        LCVLKDRNKPIIFSMARLDRVKNLTGLVEWYGKNTRLRELVNLVVVGGDRR-KESKDLEEQSEMKKMHELIETYKLNGQFRWISSQMDRVRNGELYRYIA
3660531          LCVLKDRTKPILFTMARLDRVKNLTGLVEWYAKNPRLRGLVNLVVVGGDRR-KESKDLEEQAEMKKMYELIETHNLNGQFRWISSQMNRVRNGELYRYIA
31455440         LCVLKDRTKPILFTMARLDRVKNLTGLVEWYAKNPRLRGLVNLVVVGGDRR-KESKDLEEQAEMKKMYELIETHNLNGQFRWISSQMNRVRNGELYRYIA
304651488        LCVLKDRNKPIIFTMARLDRVKNLTGLVEWYAKNPRLRELVNLVVVGGDRR-KESKDLEEQAEMKKMYELIKTHNLNGQFRWISSQMNRVRNGELYRYIA
28629438         LCVLKDRNKPIIFTMARLDRVKNLTGLVEWYAKNPRLRELVNLVVVGGDRR-KESKDLEEQAEMKKMYELIKTHNLNGQFRWISSQMNRVRNGELYRYIA
115310618        LCVLKDKKKPILFTMARLDRVKNLTGLVELYAKNPKLRELVNLVVVGGDRR-KESKDLEEQAEMKKMYSLIETYNLNGQFRWISSQMNRVRNGELYRYIA
345286417        LCVLKDKNKPIIFTMARLDRVKNPTGLVELYAKSPKLRQLVNLVIVGGDRR-KESKDLEEQAEMKKMYNLIETYKLNGQFRWISSQMNRVRNGELYRCIA
38425095         ICVLKDRKKPIIFSMARLDRVKNMTGLVEWYGKNKKLRELVNLVVVAGDRR-KESKDTEEKEEMKKMYSLIEEYNLNGQFRWISAQMNRVRNGELYRYIA
15235300         VGTLSDRSKPILFSMARLDKVKNISGLVEMYSKNTKLRELVNLVVIAGNIDVNKSKDREEIVEIEKMHNLMKNYKLDGQFRWITAQTNRARNGELYRYIA
297814081        VGTLSDRSKPILFSMARLDKVKNISGLVEMYSKNTKLRELVNLVVISGNIDVNKSNDREEIAEIEKMDNLVKSYKLDGQFRWITAQTNRARNGELYRYIA
6682841          VGTLSDRSKPIVFSMARLDHVKNMTGLVECYGKNSRLRELVNLVVVAGYIDVNKSKDREEIAEIEKMHELMKTYKLDGQFRWIAAQTNRARNGELYRYIA
```


255564236 IGTLKDKSKPLIFSMARLDRVKNITGLVEMYGKNAKLRELVNLVIVAGYIDVKKSKDREEIAEIEKMHDLMKKYNLEGQFRWITAQTNRARNGELYRYIA
313770763 IGTLTDKSKPIIFSMARLDRVKNISGLVECYGKNARLRELVNLVVVAGYIDVKKSNDREEILEIEKMHELMKKYKLDGQFRWLTAQTNRARNGELYRYIA
3915037 IGSLTDRSKPIIFSMARLDRVKNITGLVESYAKNSKLRELVNLVVVAGYIDVKKSSDREEIEEIEKMHDLMKQYNLNGEFRWITAQTNRARNGELYRYIA
225437428 IGMLNDRSKPIIFSMARLDQVKNITGLVECYAKNAKLREMANLVVVAGYNDVKKSNDREEIVEIEKMHDLMKEYNLHGQFRWMSSQTNRARNGELYRYIA
162458268 IGHLDDRSKPILFSMARLDRVKNITGLVEAFAKCAKLRELVNLVVVAGYNDVNKSKDREEIAEIEKMHELIKTHNLFGQFRWISAQTNRARNGELYRYIA
242035817 IGHLDDRSKPILFSMARLDRVKNITGLVEAFAKCTKLRELVNLVVVAGYNDVKKSKDREEIAEIEKMHELIKTYNLFGQFRWISAQTNRARNGELYRYIA
115452927 IGHLDDRSKPILFSMARLDRVKNITGLVEAYAKNARLRELVNLVVVAGYNDVKKSKDREEIAEIEKMHELIKTYNLFGQFRWISAQTNRARNGELYRYIA
241896730 IGHLDDPSKPILFSMARLDRVKNMTGLVKAYSKNAKLRSLVNLVVVAGYNDVKKSKDREEIAEIEKMHELIKTYNLFGQFRWISAQTNRVRNGELYRYIA
304651490 IGSLNDQSKPIIFSMARLDRVKNITGLVECYAKNATLRELANLVVVAGYNDVKKSNDREEIAEIEKMHALMKEHNLDGQFRWISAQMNRARNGELYRYIA
29289943 IGNLNDQSKPIIFSMARLDRVKNITGLVECYAKNATLRELANLVVVAGYNDVKKSNDREEIAEIEKMHALMKEHNLDGQFRWISAQMNRARNGELYRYIA
115310620 IGTLKDASKPIIFSMARLDRVKNITGLVECYAKNAELRELANLVVVAGYNDVKKSSDREEISEIEKMHMLMKEYNLDGEFRWIAAQTNRARNGELYRYIA
4468153 IGVLNDQSKPIIFSMARLDKVKNITGLVEMYAKNAKLRELANLVVVAGYNDVKKSSDREEIAEIEKMHSLIKEYKLDGQLRWISSQTNRVRNGELYRYVA
345286419 VGFLEDPSKPIIFSMARLDRVKNISGLVELYAKNARLRELANLVVVAGYIDVKKSSDREEISEIEKMHALIKQYDLDGQLRWISAQTNRARNGELYRYIA
4468151 IGVLKDRSKPIIFSMARLDRVKNITGLVEMYAKNKKLRELTNLVVVAGYNDVKKSSDREEKDEIEKMHNLIDQYDLNGSLRWISAQSNKARNGELYRYIA
151176306 IGVLDDTSKPIIFSMARLDRVKNITGLVECYGKNAKLRELANLVVVAGYNDVKKSNDREEIAEIEKMHRLIQEYNLRGQFRWIASQTNRVRNGELYRYIC
14530225 IGVLDDTSKPIIFSMARLDRVKNITGLVECYGKNAKLRELANLVVVAGYNDVKKSNDREEIAEIEKMHRLIQEYNLRGQFRWIASQTNRVRNGELYRYIC
15239816 VGLLSDQSKPIIFSMARLDRVKNLTGLVECYAKNSKLRELANLVIVGGYIDENQSRDREEMAEIQKMHSLIEQYDLHGEFRWIAAQMNRARNGELYRYIA
297795665 VGLLSDQSKPIIFSMARLDRVKNLTGLVECYAKNSKLRELANLVIVGGYIDVNQSRDREEMAEIQKMHSLIEQYDLHGQFRWIAAQMNRARNGELYRYIA
225431790 IGMLSDRSKPIIFSMARLDRVKNITGLVECFGKSSKLRELVNLVVVAGYIDVTKSRDREETKEIEKMHDLIKKYNLHGQFRWIPAQMNRARNGELYRYIA
255551835 IGYLTDQSKPIIFSMSRLDRVKNLTGLVEWYGKSSKLRELVNLVVVGGSMDVNKSRDREEMAEIKKMHGLITEYNLAGQFRWVAAQMNRARNGELYRYIA
168009716 IG-VIDKSKPILFTMARLDKVKNLTGLVELYGKNNKLKELTNLVIVGGEINPAKSKDREEVKEIAKMHDFIKEYNLHNSFRWIRSQTNRVQNGELYRYIA
168035060 VG-IIDKDKPILFTMARLDRVKNLTGLVELYGKNEKLREMTNLVIVGGEIDPAKSKDREEVKEIEKMHSFIKQYNLHNHFRWIRSQTNRVQNGELYRYIA
168058907 IG-LIDKEKPILFSMARLDRVKNLTGLVEMYGKNQKLKEFVHLVIVGGEINPSKSKDREEVREIEKMHNLIKRYKLENNFRWIRSQTNRIRNGELYRYIA
168029793 IGSLKDRNKPILFSMARLDKVKNISGLVEMFAKNPRLRELVNLVVVAGNIQKEKSKDREEMAEIDKMHNLMKEYELDGDFRWLCAQTDRVLNGELYRYIA
94266940 LGRLAEPDKPLLFTIARLDRIKNLTGLVEAYGRDSELRQRVNLVMVASVIDPQRSQDPEEAAEIRRMHEILEQYQLYGQVRWIGKFLGKVETGEAYRLVA
94264333 LGRLAEPDKPLLFTIARLDRIKNLTGLVEAYGRDSELRQRVNLVMVASVIDPQRSQDPEEAAEIRRMHEILEQYQLYGQVRWIGKFLGKVETGEAYRLVA
297569306 LGRLENLDLPPLFTIARLDRIKNLTGLVEAYGQDEELRRRVNLIMVASVTDPERSKDAEEAAEIRKMHGIIEQYGLRGQVRWVGKFLGKAETGEAYRIMA
254416162 YGKLDDPTKRPLFSIARLDRIKNLTGLTECFGKSKALQEQCNLIFVAGKLRTEDSTDNEEKDEIVKLYRLIDEYNLHGKIRWLGVRLPKLDSGEIYRVIA
332712456 FGKLDYPNKRPIFSMARLDRIKNLTGLAECFGKSRKLQERCNLILIAGNLRTEDSSDSEEKAEIIKLYQIIEEYNLYGKIRWLGVRLSKSDSGEVYRVIA
17232477 FGKLDDPNKRPIFSMARLDRIKNLTGLAECFGQSQELQERCNLILVAGKLRIEESEDNEEKDEIVKLYRIIDEYNLHGKIRWLGVRLSKNDSGEIYRVIC
75908500 FGKLDDPNKRPIFSMARLDRIKNLTGLAECFGKSKELQEHCNLILVAGKLRIEESEDNEERDEIVKLYRIIDEYNLHGKIRWLGVRLSKNDSGEIYRVIC
186682280 FGKLDDPNKRPLFSMARLDHIKNLTGLAECYGQSKELQEHCNLILVAGKLRVEESGDNEERDEIIKLYNIIDEYNLHGKIRWLGVRLSKTDSGEIYRVIA
119512682 FGKLDDPTKRPLFSVARLDRIKNLTGLAECYGKSPELQEHCNLILVAGKLRVEDSGDNEERDEIIKLYQIIEQYNLYGKIRWLGVRLTKSDSGEIYRVIA
334117431 FGKLDDPHKRPLFSMARLDRIKNMTGLAELFGKSPELQEKCNLILVAGKLRVEETEDYEEAEEIKKLYAIIDQYNLHGKIRWLGVRLSKSLSGEIYRVIA
218440696 FGKLDNPSKRPIFSMARLDRIKNLTGLAECFGRSPALQECCNLILVAGKLTVNESSDSEEREEIEKLYRIIDEHNLYGKIRWLGVRLPKADSGEIYRVIA
159031025 FGHLDNPDKRPLFSLARLDRIKNLTGLAECFGKSEALQEQCNLILVAGKLRAEDSTDREEISEIQKLYHLIDQYNLEGKIRWLGIMLPKADAGEIYRIIA
220907171 WGKLDHPDKRPLFSMARLDRIKNLTGLAEAFGQNPELQQHCNLILIAGKLRVEETIDHEEAMEIERLYEIIDRYQLTGKMRWLGVRLSKTDSGEVYRIIA
22298591 YGYLEAPEKRPLFSMARLDRIKNLTGLAEAFGRSKALQERCNLILVAGKLRTADSSDREEIAEIEKLYQIIHQYNLHGKIRWLGIRLPKADSGEIYRIIA
158339122 FGHLDNPGKPPLFSMARLDRIKNLTGLAECFGQHPQLQDHYNLILVAGKLRTSDSVDHEEINEIERLYRIIDQYNLHGKIRWLGVRFPKQDSGEIYRVVA
284053161 FGKLADPTKPPLFSVARLDRIKNITGLVECYGQHPELQEKANLIFIAGKLRVEDSSDYEEAEEIKKMYHLIEHYNLYDKVRWLGVRLSKTDTGEMYRVIA
209523126 FGKLADPTKPPLFSVARLDRIKNITGLVECYGQHPELQEKANLIFIAGKLRVEDSSDYEEAEEIQKMYHLIEHYNLYDKVRWLGVRLSKTDTGEMYRVIA
17228554 FGYLDQPQKPPIFAVAPITSIKNLTGLAECFGRSQELQAHSNLILLTSKLNIDESTNPEEAREIEKLHNIINQYQLHGHIRWLGLRLPNQEVGEAYRLVA
75909957 FGYLEQPQKPPIFAVAPITSIKNLTGLAECFGRSQELQAHSNLILLTSKLNIDETTNPEEAREIEKLHNIINQYQLQGHIRWLGLRLPNQEVGEAYRLVA
186685043 LGHLDRPNKRPIFSVSSISSIKNLAGLAECFGQSQRLQEHCNLILLSSKLHPDEATNPEEAEEIQKLHNIIDRYHLHSKIRWLGMRIPSSSLGEAYRVVA
119510624 IGKLDNPNKRPILTVAPITSVKNLTGLAECFAKNRGLQEHCNLIFITTKLYVNQATNPKEAEEIQRLHDIINQYELHGNIRCIGMRLPSPDLGEAYRVIA
298489784 LGNIDDIHKRPIFAVAPVTSIKNLTGLVECFGKSQELQTRCNLILLTSNLSVNEATYPEEAGEIDKLHNLIKEHHLQGKIRWLGMRLPSQKIGEAYRIIA
220909283 LGHLDQPEKKPLLTIAPISSIKNLTGLVECFGRSPELQNHCNLILITSKLHPSEASHAEEETEITQLHDLINQYQLHGRVRWLGLPLPHTDLGEIYRVVA

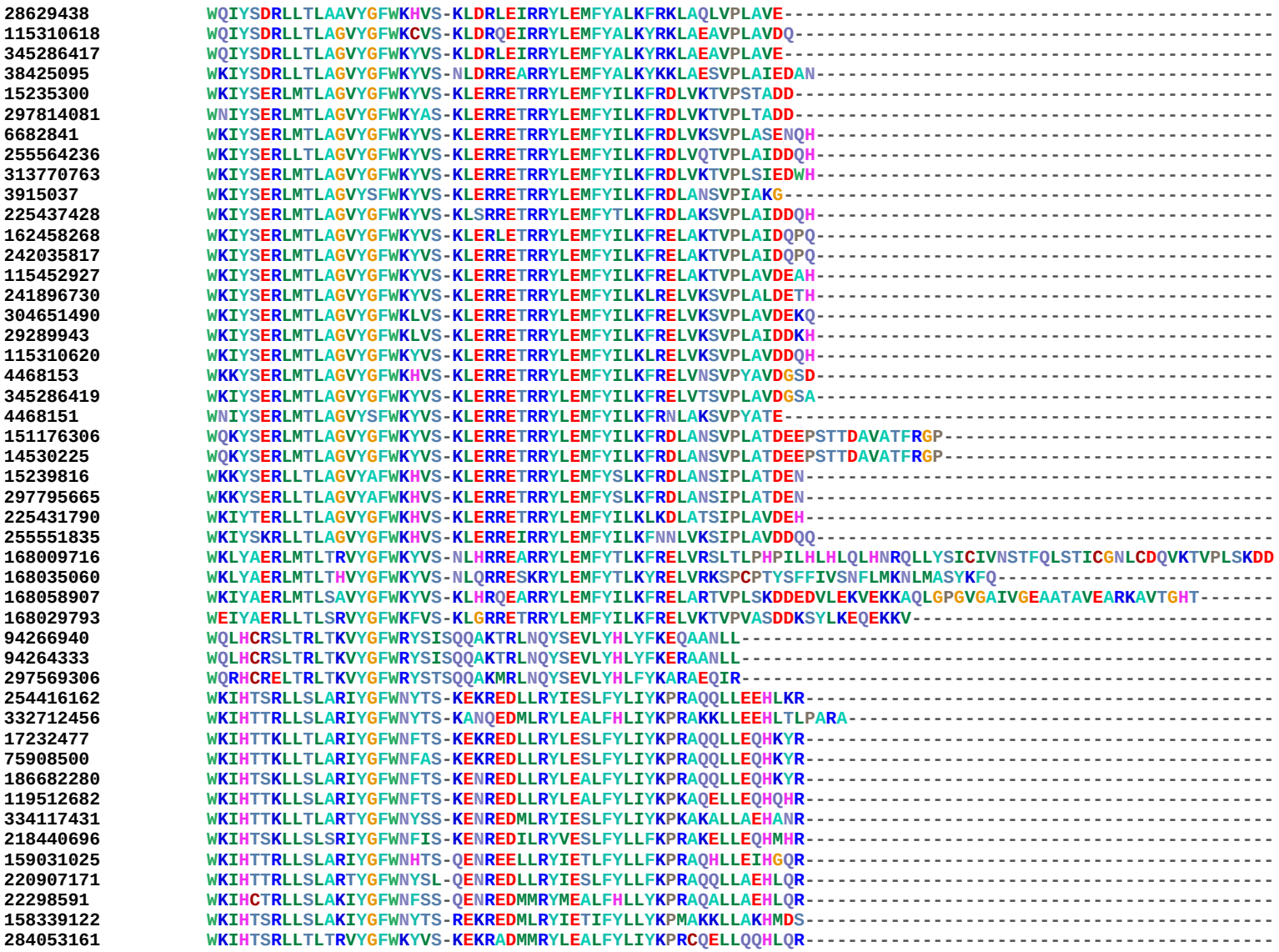

332708740 VGYLNSPNQRPILSIAPLTSIKNLSGLVECFASSKELQQKCNLILITSHVRVEDATDPEEKGEIEKLNQLIKQYNLHGKIRWIGLRLTTPDIGESYRVIA
Clustal Consensus
710 720 730 740 750 760 770 780 790 800
114331077 DKRGIFVQPALFEAFGLTIIEAMASG--------LPTFATRYGGPLEIIQHNRSGFHIDPNQGTATADLIADFLEKSHEKPLEWERLSQGALARVASRYT
30249199 DKRGIFVQPALFEAFGLTIIEAMASG--------LPTFATRYGGPLEIIQNNRSGFHIDPNQGAATADLIADFFEKNLENPQEWERISQGALDRVASRYT
82703384 DRRGIFVQPARFEAFGLTIIEAMASG--------LPVFATCYGGPREIIQHGVSGYHFDPNDGLAGASAMADFFERVAADPGFWDRISQKALQRVEARYT
198283392 DSRGCFVQPALFEAFGLTVIEAMSSG--------LPVFATRFGGPLEIIEDGISGFHIDPNNQQETAEKLADFLEAAAADIRVWETISDGALARVGAHYT
344199811 DTHGCFVQPALFEAFGLTVIEAMSSG--------LPVFATRFGGPLEIIEDGVSGFHIDPNNQQETAEKLADFLAAAAADIRVWETISDGALARVSTHYT
255021595 DGRGVFVQPALFEAFGLTVIEAMSSG--------LPVFATRFGGPLEIIEDGVSGFHIDPNDHEATAERLADFLEAARERPKYWLEISDAALARVAERYT
77166514 DSRGAFVQPALFEAFGLTVIEAMSSG--------LPTFATCYGGPLEIIQEGVSGFHIDPNHGEKAADRIADFFEHCQTEAGYWDKFSQGALRRIKNHYT
300115586 DSRGAFVQPALFEAFGLTVIEAMSSG--------LPTFATCYGGPLEIIQEGVSGFHIDPNHGEKVANRIADFFEHCQTEAGYWDRFSQGALHRIKNHYT
292493898 DSRGAFVQPALFEAFGLTVIEAMSSG--------LPTFATCYGGPLEIIQDEISGFHINPNHGEEAAGSIADFFERCQVEPEYWENLSQGALRRIRRRYT
350552412 DRRGVFVQPALFEAFGLTVIEAMVSG--------LPTFATHYGGPLEIIEHGQSGFHIDPMRGDQASAQLLAFLRECEQDPDYWVRISHGGMERVERHYT
220933364 DRRGVFVQPALFEAFGLTVIEAMASG--------LPTFATRYGGPLEIIEDGVSGYHIDPNHGEQAARILMEFLERCASDPDHWQQISRSAIRRVEQRYT
350574485 DHQGAFVQPALFEAFGLTVVEAMSCG--------LPTFATCYGGPSEIIEHGLSGFHIDPNHGDQAAALILEFFDACSQNPAHWQTFSTAAMARVQERYT
95929190 DQKGVFVQPALFEAFGLTVIEAMATG--------LPIFATQYGGPLEIIVDGKSGFHIDPNDNEEMAEKICTFFERAANHPQYWKVISDACITRVEENYT
298528445 DSRGAFVQPALFEAFGLTVVEAMNSG--------LPTFATIFGGPLEIIEDGKSGFHIDPTHGDEAAGLMANFFSRCRADASYWDTISNNSIKRVEEKYN
297602308 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIIDGVSGFHVNPINGREAGIKIADFFQKCKEDPSYWNKVSTAGLQRIYEWQR
115457664 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIIDGVSGFHVNPINDREAGIKIADFFQKCKEDPSYWNKVSTAGLQRICECYT
326504012 DSKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDGVSGFHINPMNGREAGTKIADFFQKCKEDPSYWNKMSTAGLQRIYECYT
115450038 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDEVSGFHINPLNGKEASDKIADFFQKCKEDLIYWSKMSTAGLQRIYECYT
326531526 DTRGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVNEVSGFHINPLNGKESSDKIAAFFQKCKEDPTYWNKMSTAGLQRIYECYT
242063616 DTRGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDEVSGFHINPLDGKEASNKIANFFQKCNEDPMYWNRMSTAGLQRIYECYT
42568160 DTRGAFVQPAHYEAFGLTVIEAMSCG--------LVTFATNQGGPAEIIVDGVSGFHIDPSNGEESSDKIADFFEKSGMDPDYWNMFSNEGLQRINECYT
297805240 DTRGAFVQPAHYEAFGLTVIEAMSCG--------LVTFATNQGGPAEIIVDGVSGFHIDPSNGEESSDKIADFFEKSGTDLDYWNMFSTEGLQRINECYT
313770765 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPSEIIVDGISGFHIDPKNGDESSNIIADFFEKCKVDPGHWNKYSLEGLKRINECYT
313770767 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDGISGFHIDPQNGDESSNIIADFFEKCKVDPGYWNKFAAEGLKRINECYT
255584097 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIIDGVSGFLIDPNNGDESSNKIADFFEKCKIDAEYWNKFSEDGLKRINECYT
225466221 DTKGAFVQPAIYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDGVSGFHIDPNIGDESSNKIADFFEKCRDDSDHWNKISKAGLQRINECYT
313770769 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEILVDGISGFHIDPNNGDESSNKIADFFEKCKTDAEYWNKMSAAGLQRIYECYT
313770771 DTKGAFIQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDGISGFHIDPNNGDESSNKIADFVEKCKTDAEYWNKMSATGLQRIYECYT
255570671 DTKGAFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDGVSGFHIDPNNGNESSNKIADFFEKCKADPECWNKMSAAGLQRIHECYT
15219457 DTKGVFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDGVSGFHIDPNNGDESVTKIGDFFSKCRSDGLYWDNISKGGLKRIYECYT
297842089 DTKGVFVQPALYEAFGLTVIEAMNCG--------LPTFATNQGGPAEIIVDGVSGFHIDPNNGDESVTKIGDFFSKCSSDGLYWDNISKAGLKRIYESYT
115466896 DTKGVFVQPAFYEAFGLTVIEAMTCG--------LPTIATCHGGPAEIIVDGVSGLHIDPYHSDKAADILVNFFEKCKQDSTYWDNISQGGLQRIYEKYT
17980243 DTKGVFVQPAFYEAFGLTVIESMTCG--------LPTIATCHGGPAEIIVDGVSGLHIDPYHSNKAADILVNFFEKCKEDPTYWDKISQGGLKRIYEKYT
162460741 DTKGAFVQPAFYEAFGLTVIESMTCG--------LPTIATCHGGPAEIIVDGVSGLHIDPYHSDKAADILVNFFDKCKADPSYWDEISQGGLQRIYEKYT
239984690 DTKGAFVQPAFYEAFGLTVIEAMTCG--------LPTIATCHGGPAEIIVDGVSGLHIDPYHSDKAADILVNFFEKSTADPSYWDKISQGGLKRIYEKYT
162460681 DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATAYGGPAEIIVHGVSGYHIDPYQGDKASALLVDFFDKCQAEPSHWSKISQGGLQRIEEKYT
242035533 DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATAYGGPAEIIVHGVSGFHIDPYQGDKASALLVDFFEKCQTDSSHWNKISQGGLQRIEEKYT
62865493 DTKGAFVQPAFYEAFGLTVVESMTCG--------LPTFATAYGGPAEIIVHGVSGFHIDPYQGDKASALLVEFFEKCQQDHSHWTKISQGGLQRIEEKYT
17980241 DTRGAFVQPAFYEAFGLTVVESMSCG--------LPTFATAYGGPAEIIVHGVSGFHIDPYQGDKASALLVEFFEKCQQDPAHWTKISQGGLQRIEEKYT
115453437 DTKGAFVQPAFYEAFGLTVVESMTCG--------LPTFATAYGGPAEIIVNGVSGFHIDPYQGDKASALLVEFFEKCQEDPSHWTKISQGGLQRIEEKYT
115473359 DMRGAFVQPALYEAFGLTVIEAMTCG--------LPTFATAYGGPAEIIVHGVSGYHIDPYQNDKASALLVEFFEKCQEDPNHWIKISQGGLQRIEEKYT
3980298 DMKGAFVQAAFYEAFGLTVIEAMTCG--------LPTFATAYGGPAEIIVNGVSGYHIDPYQNDKASALLVDFFGKCQEDPSHWNKISQGGLQRIEEKYT

```
15242073        DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCKGGPAEIIVHGKSGFHIDPYHGDQAADTLADFFTKCKEDPSHWDEISKGGLQRIEEKYT
297812265       DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCKGGPAEIIVHGKSGFHIDPYHGDQAANTLADFFTKCKEDPSHWDEISKGGLQRIEEKYT
22331535        DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVHGKSGFHIDPYHGDKAAESLADFFTKCKHDPSHWDQISLGGLERIQEKYT
297818772       DTKGAFVQPALYEAFGLTVVEAMTSG--------LPTFATCNGGPAEIIVHGKSGFHIDPYHGDQAAETLADFFTKCKHDPSHWDQISLGGLERIQEKYT
313770759       DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVNGKSGFHIDPYHGEKAAELLVDFFEKCKVDPAHWDKISHGGLQRIQEKYT
313770761       DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVHGKSGFHIDPYHGVQAAELLVDFFEKCKADPSYWDKISQGGLQRIQEKYT
255550319       DTKGVFVQPALYEAFGLTVVESMSCG--------LPTFATCNGGPAEIIVHGKSGFNIDPYHGDQAAELLVEFFEKCKADPCVWDEISKGGLQRIQEKYT
324984223       DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVHGKSGFNIDPYHGDQAADILVDFFEKCKKDPSHWDKISQGGLKRIEEKYT
258489633       DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVHGKSGFNIDPYHGDQAADILVDFFEKCKKDPSHWDKISQGGLKRIEEKYT
324984221       DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVHGKSGFNIDPYHGDQAADILVDFFEKCKKDPSHWDKISQGGLKRIEEKYT
324984229       DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVHGKSGFNIDPYHGDQAADILVDFFEKCKKDPSHWDKISQGGLKRIEEKYT
6683114         DTKGAFVQPALYEAFGLTVVEAMTCG--------LPTFATCKGGPAEIIVNGKSGYHIDPYHGEQAAEILVDFFEKCKADPSYWDKISLGGLKRIEEKYT
4584690         DTKGAFVQPAVYEAFGLTVVEAMATG--------LPTFATLNGGPAEIIVHGKSGFHIDPYHGDRAADLLVEFFEKVKVDPSHWDKISQGGLQRIEEKYT
3915046         DTKGAFVQPAVYEAFGLTVVEAMATG--------LPTFATLNGGPAEIIVHGKSGFHIDPYHGDRAADLLVEFFEKVKADPSHWDKISQGGLQRIEEKYT
3766299         DTKGAFVQPAVYEAFGLTVVEAMATG--------LPTFATLNGGPAEIIVHGKSGFHIDPYHGDRAADLLVEFFEKVKTDPSHWDKISQGGLQRIEEKYT
3377764         DTKGAFVQPAVYEAFGLTVVEAMATG--------LPTFATLNGGPAEIIVHGKSGFHIDPYHGDRAADLLVEFFEKVKTDPSHWDKISQGGLQRIEEKYT
225444613       DTKGVFVQPAFYEAFGLTVVEAMTCG--------LPTFATCNGGPAEIIVHGKSGFHIDPYHGDKAAELLANFFEKCKADPTHWEKISKAGLKRIEEKYT
3660531         DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATNHGGPAEIIVHGKSGFHIDPYHGEQAADLLADFFEKCKKEPSHWETISTGGLKRIQEKYT
31455440        DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATNHGGPAEIIVHGKSGFHIDPYHGEQAADLLADFFEKCKKEPSHWETISTGGLKRIQEKYT
304651488       DTRGAFVQPAFYEAFGLTVVEAMSCG--------LPTFATNQGGPAEIIVHGKSGFQIDPYHGEQAADLLAEFFEKCKVDPSHWEAISKGGLKRIQEKYT
28629438        DTRGAFVQPAFYEAFGLTVVEAMSCG--------LPTFATNQGGPAEIIVHGKSGFQIDPYHGEQAADLLADFFEKCKVDPSHWEAISEGGLKRIQEKYT
115310618       DTRGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATNHGGPAEIIIHGKSGFHIDPYHGEQVSELLANFFERCKKEPSYWDTISAGGLKRIQEKYT
345286417       DTKGAFVQPAFYEAFGLTVVEAMTRG--------LPTFATLHGGPAEIIVDGKSGFHIDPYNGEQVAETLVSFFEKCNKDPSHWEAISTGGLKRIQEKYT
38425095        DTRGAFVQPAYYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIVHGKSGFHIDPYHGDKAADLLVDFFEKSTADPSYWENISKGGLQRIEEKYT
15235300        DTRGAFAQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIEHGLSGFHIDPYHPEQAGNIMADFFERCKEDPNHWKKVSDAGLQRIYERYT
297814081       DTRGAFAQSIRLLAMKTCILRGFWAYGSGSDDLRASDFRHLSGGPAEIIEHGLSGFHIDPYHPEQAGNIMADFFERGREDPNHWKKVSDAGLQRIYERYT
6682841         DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIEHGASGFHIDPYHPDQAAELMADFFGKCKENPSHWKKISDGGLKRIYERYT
255564236       DTKGAFVQPAFYEAFGLTVVEAMTSG--------LPTFATCHGGPAEIIVDGVSGFHIDPYHPDQAAAIMADFFQQCKEDPSHWNKISDAGLQRIYERYT
313770763       DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIEHGVSGFHMDPYYPDQAAEFMADFFEKCKDDPSYWKKISDAGLQRIYERYT
3915037         DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATNHGGPAEIIEHGVSGFHIDPYHPDQASELLVDFFQRCKEDPNHWNKVSDGGLQRIYERYT
225437428       DTRGIFVQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIENGVSGFHIDPYHPDQVATTMVDFFEKCKEDSSHWNKISDAGLQRIYERYT
162458268       DTHGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATLHGGPAEIIEHGVSGFHIDPYHPEQAANLMADFFDRCKQDPDHWVNISGAGLQRIYEKYT
242035817       DTHGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATLHGGPAEIIEHGISGFHIDPYHPEQAANLMADFFERCKQDPNHWVKISEAGLKRIYEKYT
115452927       DTHGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATVHGGPAEIIEHGISGFHIDPYHPDQAANLIADFFEQCKQDPNHWVEVSNRGLQRIYEKYT
241896730       DTHGAFVQPALYEAFGLTVVEAMTCG--------LPTFATLHGGPAEIIEHGISGFHIDPYHPDQAATLMADFFGQCKQDPNHWVKISDKGLQRIYEKYT
304651490       DKRGIFVQPAYYEAFGLTVVEAMTCG--------LPTFATCHGGPMEIIQDGVSGYHIDPYHPNKAAELMVEFFQRCEQNPTHWENISASGLQRILDRYT
29289943        DKRGIFVQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPMEIIQDGVSGYHIDPYHPNKAAELMVEFFQRCEQNPTHWENISASGLQRILDRYT
115310620       DKRGIFVQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPKEIIEDGISGFHIDPYHPDKDSAAMVNFFQRCKEDPKYWEKISRGGLQRIYERYT
4468153         DTRGIFVQPAFYEAFGLTVVEAMTCG--------LPTFATSHGGPMEIIEDRISGFHIDPYHPEKAADLMADFFGKCNEDPSYWVKISEAALRRIQERYT
345286419       DKRGIFVQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPLEIIEDGISGFHIDPYHPDKSALLMADFFEKRNEDPSYWVKISEAALRRIQERYT
4468151         DKRGIFVQPAFYEAFGLTVVEAMTCG--------LPAFATLHGGPHEIIEDGVSGFHIDPYHAEKAATRMADFFAKCDDDPSYWVKISEQALQRIRECYT
151176306       DKGGIFAQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIEDGVSGFHIDPYHADQAAEKMTEFFVKCREDPNYWTKISAGGLLRIKERYT
14530225        DKGGIFAQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIEDGVSGFHIDPYHADQAEK-MTEFFVKCREDPNYWTKISAGGLLRIKERYT
15239816        DTKGVFVQPAFYEAFGLTVVESMTCA--------LPTFATCHGGPAEIIENGVSGFHIDPYHPDQVAATLVSFFETCNTNPNHWVKISEGGLKRIYERYT
297795665       DTKGVFVQPAFYEAFGLTVVESMTCG--------LPTFATCHGGPAEIIENGVSGFHIDPYHPDQVAATLVSFFETCNTNPNHWVKISEGGLKRIYERYT
225431790       DTKGAFVQPAFYEAFGLTVVEAMTCG--------LPTFATCHGGPAEIIEDGLSGFHIDPYHPDQVALRLADFFERCQKDPSYWDEISNGGLKRIYERYT
255551835       DAKGVFVQPAFYEAFGLTVIEAMTCG--------LPTFATCHGGPAEIIEHGTCGFHIDPHHPDQAASLLINFFERCKEDPSYWNTISDGGLKRIYERYT
```

```
168009716          EAGGVFVQPALYEGFGLTVVEAMTCG--------LPTFATLHGGPAEIIEHGISGFHIDPYHPDEVADELVTFFEKVKSDSSFWTKISEAALQRIYSSFT
168035060          DAGGVFVQPALYEGFGLTVVEAMTCG--------LPTFATMHGGPAEIIVNGISGFHIDPYHPEGVAEVLVSFFEKVKTDPGVWTRISEAALQRIYSNFT
168058907          DSQGAFVQPALYEGFGLTVVEAMTSG--------LPTFATSHGGPAEIIEHGISGYHIDPYYPDEAAEQIVAFFEKCKNEPGLWNKVSEAGLQRIYSSYT
168029793          DSHGAFVQPALYEGFGLTVIEAMTCG--------LPTFATCHGGPKEIVVSDVSGFHIDPFHPESASKIIVDFFERCTKEKDYWTKLSDGGLERIRTKYT
94266940           DRGGVFVQPALFEAFGLTILEAMHSG--------LPVFATQFGGPLEIIEHEHSGFLINPTDPQAMTARLNEFFAACQADPRHWQGFSQRGLERARSRFT
94264333           DRGGVFVQPALFEAFGLTILEAMHSG--------LPVFATQFGGPLEIIEHEHSGFLINPTDPQAMTARLNEFFAACQADSRHWQGFSQRGLERARSRFT
297569306          DRRGVFVQPALFEAFGLTILEAMHSG--------LPVFATQFGGPLEIIEHEKSGFLINPTDQPAMTARLREFFHHCEENPRYWQGFSQRALERARERFT
254416162          DHRGIFVQPALFEAFGLTILEAMISG--------LPTFGTQFGGPLEIIQDKVNGFLINPTNLEETAQKILEFLSKCEQNPDYWLEISNRGMERVYSTYT
332712456          DRHGIFVQPALFEAFGLTILESMISG--------LPTFGTQFGGPLEIIQDKVNGILINPTNQEEMAQKILDFVTKCEENPQYWEEISNQGIERVYSTYT
17232477           DRQGIFVQPALFEAFGLTILESMISG--------LPTFATQFGGPLEIIQDKINGFYINPTHLEETATKILDFVTKCEQNPNYWNIISEKAIDRVYSTYT
75908500           DRQGIFVQPALFEAFGLTILESMISG--------LPTFATQFGGPLEIIQDQINGFYINPTHLEETATKILDFVTKCEHNPNYWKIISEKAIDRVYSTYT
186682280          DRQGIFVQPALFEAFGLTILESMVSG--------LPTFATQFGGPLEIIQDKVNGFLINPTNLDETATKIVDFITKCEQNPNYWNEISQRGIDRVYSTYT
119512682          DRKGIFVQPALFEAFGLTILESMVSG--------IPTFATQFGGPLEIIQDKVNGFYINPTNLEETAEKILEFVTKCEQSSHYWDAVSEEAIKRVLTTYT
334117431          DTQGIFVQPALFEAFGLTILEAMITG--------IPTFGTQFGGPLEIIQDGVNGFYINPTNHEDTAQKILDFLSKCEHNPNYWYEISTRGIDRVYSTYT
218440696          DRRGVFVQPALFEAFGLTILEAMISG--------LPTFATQFGGPLEIIQDKVNGFYINPTNLEETADKILEFVTKCDHNPDHWIQLSNKAMERVYSTYT
159031025          DRQGIFVQPALFEAFGLTVLEAMITG--------LPIFATEFGGPREIIQHGANGFLINPTHPEETATMILDFLAKCRQDPDYWREISEQAIQRVYSHYT
220907171          DHQGIFVQPALFEAFGLTILEAMITG--------LPTFATQFGGPLEIIQEGVNGFLINPTQPEATAAKILQYVRQCEDNPQTWQSISERAIERVYSTYT
22298591           DRQGIFVQPALFEAFGLTILEAMISG--------LPTFGTRFGGPLEIIQDGVNGFYINPTHLEEMAETIVRFLEACDRDPQEWQRISKAGIERVYSTYT
158339122          DHKGIFVQPALFEAFGLTVLEAMISG--------LPTFATRFGGPLEIIRDGIDGFYINPTHNEEIATKLLEFAKECATNPDYWQQISEQAIERVYTTYT
284053161          DHHGIFVQPALFEAFGLTILEAMISG--------LPTFATQFGGPLEIIKDKINGFYINPTNYDETAAKLDEFLVRCQYNIGFWNEISQRGMDRVYSSYT
209523126          DHHGVFVQPALFEAFGLTILEAMISG--------LPTFATQFGGPLEIIKDKINGFYINPTNYDETAAKLDEFLLRCQYNIGFWNEISQRGMDRVYSSYT
17228554           DYRGIYIHFARFEAFGRSILEAMISG--------LPTFATKFGGSLEIMEDQNNGFRINPTDLEGTAEKILAFFQECDTHPEHWQEVSQWMSQRIHQKYN
75909957           DYRGIYLHFARFEAFGRSILEAMISG--------LPTFATKFGGSLEILEDQNNGFRINPTDLEGTAEKILAFFQECDTHPEHWQEVSQWMSQRIHQKYN
186685043          DCQGISVHFARFESFGRSILEAMISG--------LPTFATQFGGSLEIIENQEEEFNVNPTDLVETAKKILDFFEKCNTHPEHWQEVSEWMSQRVHNRYN
119510624          DAQGIYVHFARFESFGRSILEAMVSG--------LPTFVTKFGGAVEIIQDQEETFHINPTDFKATAHQILNFIDQCETQPERWTEVSQMMSQRVINKYN
298489784          DYQGIYIHFALYEAFGRSILEAMISG--------LPTFATKFGGSSEILEDLQTGFHLNPTNLERTAKTILNFLDKCDANPEYWQETSQWMIQRIRHKYN
220909283          DQRGFFVHFARFEAFGQTILEAMISG--------LPAFATQFGGCLEIIQDGENGFHINPTDPEGTVKKILAFIHACAADPTYWQGISERAIERVQQQYN
332708740          DLGGFLVHPARFEAFGLTVLEAMISG--------LPTFATQFGGPSEIIQNGDNGFLINPTDLQDTAEKIQQFISKCEHTPDYWQKISQAGIKRVRDKYN
Clustal Consensus  :  *   :     .:   ::..:             :    **  **:       .:*        :  :          *   :     *

                          810       820       830       840       850       860       870       880       890       900
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
114331077          WKLYAERMMTLSRIYSFWKFVS-GLEREETDLYLNMFYHLQFRPLANRLAGDA-------------------------------------------------
30249199           WKLYAERMMTLSRIYGFWKFVS-GLEREETDRYLNMFYHLQFRPLANRLAHEI-------------------------------------------------
82703384           WRLYAERMMTLSRIYGFWKFVS-KLEHEETARYLNMFYHLQFRPMAQALPQ---------------------------------------------------
198283392          WGNYAAQMMTLARIFGFWRFML-KADRHAARRYLQMFQHLQWRPLAHAVPLGES-----------------------------------------------
344199811          WGNYATQMMTLARIFGFWRFML-KTDHHAARRYLQMFQHLQWRPLAHAVPLGES-----------------------------------------------
255021595          WERYAERLMTIARIFGFWRFVL-DRESQVMERYLQMFRHLQWRPLAHAVPME-------------------------------------------------
77166514           WELYAERMMTLSRIYGFWKYVT-NLERAERRRYLEMFYNLQFRPLAQQIEH---------------------------------------------------
300115586          WELYAERMMTLSRIYGFWKYVT-NLEREERRRYLEMFYNLQFRPLAQQMEH---------------------------------------------------
292493898          WDLYAERMMTLSRIYGFWKYVT-NLEREESRRYLEMFYNLQFRPLAQQMEH---------------------------------------------------
350552412          WSLYAQRMMTLSRIYGFWKYVT-NLERAEARRYLEMFYALQYRPLARSLTGE-------------------------------------------------
220933364          WKLYAERMMTLSRIYGFWKYVT-NLERAETRRYLEMFHALQYRPLALLK-----------------------------------------------------
350574485          WRRYAERMMTLSRVYGFWKYVT-DLERAETSRYLEMFYTLKLRPLAKAIQG---------------------------------------------------
95929190           WSLYARRLLTLSRVYGFWKYVS-NLEREETRRYLEMFHGLMFRNLASETQ----------------------------------------------------
298528445          WRLYAQRLLSFSRIYGFWKYVS-NLERDETRRYLDMFYSLKMRSLSH-------------------------------------------------------
297602308          LWQEQGINRLGKPQQAWRLKAA-DSDAKDLNQVSRAEASSF-------------------------------------------------------------
```


115457664 WKIYATRVLNMGSTYSFWKTLN-KEERQAKQRYLQIFYNVQYRNLAKAMARAGDQQARQTTTG---VAPSEIVVRPKERKPQTRMQRILTRLAGQKPPVS
326504012 WKIYATKVLNMGSMYGFWRTLN-KEERVAKQRYMQMFYNLQYRNLVKTVPRVGEQPPRPAASTGAVAERNQIVARPRERKPQGRVQRMMTSLLGPKPPTY
115450038 WQIYATKVLNMASIYGFWRTLD-KEERQAKQHYLHMFYNLQFRKLAKNVPTLGEQPAQ----PTESAEPNRIIPRPKERQVCPFLRNLLKKETGNN----
326531526 WQIYATKVLNMGSMYGFWRTLN-KEERQAKQLYLQMFYNLLFRQLVKTVPKLGEQPAQ----PTTAP--ARIAPRPRERRPQTRIQRIATSLLGPVLPTS
242063616 WQIYATKVLNMGSMYGFWRTLD-KEEKQAKQQYLQMFYNLHFRKLANAVPKVGEQPEQ----ATAVPLPDRSAPRPKERQVCPLLRNLLKIKWGSN----
42568160 WKIYANKVINMGSTYSYWRHLN-KDQKLAKQRYIHSFYNLQYRNLVKTIPILSDIPEPPPLPPKPLVKPSASKGSKRTQPRLSFRLFGA----------
297805240 WKIYANKVINMGSTYSYWRYLN-KDQKLAKQRYIHSFYNLQYRNLVKNIQILSDIPQPPPPPPKPLVKPSASKGSRRTQSRLSFRLFGA----------
313770765 WKIYANKLLNMGNVYSFWRQLN-KEQKLAKQRYIQLFFNLKFRELVQSVPIPTEEAQTPASEPTARTQSSAR----------
313770767 WKIYAKKLLNMGNMYSFWRQLN-KEQKLAKQRYIQMLYNLQFRRLVVWSLSCNQEAWIRTAIILCHLTLLNMFIHNVFYLLLSEFNFSVYDH--------
255584097 WKIYANKVLNMGCIYTYWRQMN-KEQKQAKRRYIQLFYNLQLRKLVKNVPIPTEEAQQQPAPKPVNKASSRQVP----------
225466221 WKIYANKVLNMGCVFSFWRQLN-TEHKQAKQKYIHMFYTLQFRNLVKNIPIPASEVQPPVSRAITKHHHQALTVSTGKVKPKSQLTDYPASAFSQVKNDL
313770769 WKIYANKVLNMGSVYGFWRQTN-KEQKLAKQRYIEAFYNLQFNNLVGYCGQLVL----------
313770771 WKIYANKVLNMGSVYGFWRQMN-KEQKLLKQRYIEAFYNLQFRNLVGYFRQLVT----------
255570671 WKIYANKVLNMGSVYGFWRQLN-KEQKHAKQRYIETFYNLHFRNLVKNVPIASVGPQKQPSSSAGTSKTQEPSPPATTKSSQSQPTPKPKRKEESQKEKP
15219457 WKIYAEKLLKMGSLYGFWRQVN-EDQKKAKKRYIEMLYNLQFKQLTKKVTIPEDKPLPLRLASLRNLLPKKTTNLGAGSKQKEVTETEKTKQKSKDGQEQ
297842089 WKIYAEKLLKMGSIYGFWRQVN-EDQKKAKQRYIELLYNLQFKQLTKKVTIPEDKPLPLRLASLRNLLPKKTTPLGAGSKQKEVTEAEKTKQKSKDGQEQ
115466896 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYIEMFYALKYRSLASAVPLAVDGESTSK----------
17980243 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYLEMFYALKYRSLASAVPLAVDGDSAAN----------
162460741 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYIEMFYALKYRSLASQVPLSFD----------
239984690 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYLEMFYALKYRSLAAAVPLAVDGESSGN----------
162460681 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYLEMLYALKYRTMASTVPLAVEGEPSSK----------
242035533 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYLEMLYALKYRTMASTVPLAVEGEPSSK----------
62865493 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYLEMLYALKYRTMASTVPLAVDGEPSSK----------
17980241 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYLEMLYALKYRKMASTVPLAVEGEPSNK----------
115453437 WKLYSERLMTLTGVYGFWKYVS-NLERRETRRYLEMLYALKYRTMASTVPLAVEGEPSNK----------
115473359 WKLYSERLMTLSGVYGFWKYVT-NLDRRETRRYLEMLYALKYRKMATTVPLAIEGEASTK----------
3980298 WKLYSERLMTLSGVYGFWKYVS-NLDRRETRRTLKCSMPSSTAKWLQLSHWLLRARPRANDLSLPEK----------
15242073 WQIYSQRLLTLTGVYGFWKHVS-NLDRLEARRYLEMFYALKYRPLAQAVPLAQDD----------
297812265 WQIYSQRLLTLTGVYGFWKHVS-NLDHLEARRYLEMFYALKYRPLAQAVPLAQDD----------
22331535 WQIYSQRLLTLTGVYGFWKHVS-NLDRLESRRYLEMFYALKYRPLAQAVPLAHEE----------
297818772 WQIYSERLLTLTGVYGFWKHVS-NLDRLESRRYLEMFYALKYRPLAQAVPLAQEE----------
313770759 WQIYSQRLLTLTGVYGFWKHVS-NLDRLESRRYMEMFYALKYRKLAESVPLTKE----------
313770761 WKIYSQRLLTLTGVYGFWKHVS-NLDHRESRRYLEMFYALKYRKLADSVPLTIE----------
255550319 WQIYSQRLLTLTGVYGFWKHVS-KLDRRESRRYLEMFYALKYKKLADSVPLTVE----------
324984223 WKIYSERLLTLTGVYGFWKHVS-NLERRESRRYLEMFYALKYRKLAESVPLAEE----------
258489633 WKIYSERLLTLTGVYGFWKHVS-NLERRESRRYLEMFYALKYRKLAESVPLAEE----------
324984221 WKIYSERLLTLTGVYGFWKHVS-NLERRESRRYLEMFYALKYRKLAESVPLAEE----------
324984229 WKIYSERLLTLTGVYGFWKHVS-NLERRESRRYLEMFYALKYRKLAESVPLAEE----------
6683114 WKIYSQRLLTLTGVYGFWKHVS-NLDRLESRRYLEMFYALKYRKLAESVPLAVE----------
4584690 WTIYSQRLLTLTGVYGFWKHVS-NLDRLESRRYLEMFYALKYRKLAESVPLAVE----------
3915046 WTIYSQRLLTLTGVYGFWKHVS-NLDRLESRRYLEMFYALKYRKLAESVPLAVE----------
3766299 WQIYSQRLLTLTGVYGFWKHVS-NLDRLESRRYLEMFYALKYRKLAESVPLAVEE----------
3377764 WQIYSQRLLTLTGVYGFWKHVS-NLDRLESRRYLEMFYALKYRKLAESVPLAVEE----------
225444613 WKIYSERLLTLAGVYGFWKYVS-NLDRRETRRYLEMFYALKYRKLAQSVPLAVEE----------
3660531 WQIYSERLLTLAAVYGFWKHVS-KLDRLEIRRYLEMFYALKYRKMAEAVPLAAE----------
31455440 WQIYSERLLTLAAVYGFWKHVS-KLDRLEIRRYLEMFYALKYRKMAEAVPLAAE----------
304651488 WQIYSDRLLTLAAVYGFWKHVS-KLDRLEIRRYLEMFYALKFRKLAELVPLAVE----------


28629438 WQIYSDRLLTLAAVYGFWKHVS-KLDRLEIRRYLEMFYALKFRKLAQLVPLAVE
115310618 WQIYSDRLLTLAGVYGFWKCVS-KLDRQEIRRYLEMFYALKYRKLAEAVPLAVDQ
345286417 WQIYSDRLLTLAGVYGFWKYVS-KLDRLEIRRYLEMFYALKYRKLAEAVPLAVE
38425095 WKIYSDRLLTLAGVYGFWKYVS-NLDRREARRYLEMFYALKYKKLAESVPLAIEDAN
15235300 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRDLVKTVPSTADD
297814081 WNIYSERLMTLAGVYGFWKYAS-KLERRETRRYLEMFYILKFRDLVKTVPLTADD
6682841 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRDLVKSVPLASENQH
255564236 WKIYSERLLTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRDLVQTVPLAIDDQH
313770763 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRDLVKTVPLSIEDWH
3915037 WKIYSERLMTLAGVYSFWKYVS-KLERRETRRYLEMFYILKFRDLANSVPIAKG
225437428 WKIYSERLMTLAGVYGFWKYVS-KLSRRETRRYLEMFYTLKFRDLAKSVPLAIDDQH
162458268 WKIYSERLMTLAGVYGFWKYVS-KLERLETRRYLEMFYILKFRELAKTVPLAIDQPQ
242035817 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRELAKTVPLAIDQPQ
115452927 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRELAKTVPLAVDEAH
241896730 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKLRELVKSVPLALDETH
304651490 WKIYSERLMTLAGVYGFWKLVS-KLERRETRRYLEMFYILKFRELVKSVPLAVDEKQ
29289943 WKIYSERLMTLAGVYGFWKLVS-KLERRETRRYLEMFYILKFRELVKSVPLAIDDKH
115310620 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKLRELVKSVPLAVDDQH
4468153 WKKYSERLMTLAGVYGFWKHVS-KLERRETRRYLEMFYILKFRELVNSVPYAVDGSD
345286419 WKIYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRELVTSVPLAVDGSA
4468151 WNIYSERLMTLAGVYSFWKYVS-KLERRETRRYLEMFYILKFRNLAKSVPYATE
151176306 WQKYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRDLANSVPLATDEEPSTTDAVATFRGP
14530225 WQKYSERLMTLAGVYGFWKYVS-KLERRETRRYLEMFYILKFRDLANSVPLATDEEPSTTDAVATFRGP
15239816 WKKYSERLLTLAGVYAFWKHVS-KLERRETRRYLEMFYSLKFRDLANSIPLATDEN
297795665 WKKYSERLLTLAGVYAFWKHVS-KLERRETRRYLEMFYSLKFRDLANSIPLATDEN
225431790 WKIYTERLLTLAGVYGFWKHVS-KLERRETRRYLEMFYILKLKDLATSIPLAVDEH
255551835 WKIYSKRLLTLAGVYGFWKHVS-KLERREIRRYLEMFYILKFNNLVKSIPLAVDDQQ
168009716 WKLYAERLMTLTRVYGFWKYVS-NLHRREARRYLEMFYTLKFRELVRSLTLPHPILHLHLQLHNRQLLYSICIVNSTFQLSTICGNLCDQVKTVPLSKDD
168035060 WKLYAERLMTLTHVYGFWKYVS-NLQRRESKRYLEMFYTLKYRELVRKSPCPTYSFFIVSNFLMKNLMASYKFQ
168058907 WKIYAERLMTLSAVYGFWKYVS-KLHRQEARRYLEMFYILKFRELARTVPLSKDDEDVLEKVEKKAQLGPGVGAIVGEAATAVEARKAVTGHT
168029793 WEIYAERLLTLSRVYGFWKFVS-KLGRRETRRYLEMFYILKFRELVKTVPVASDDKSYLKEQEKKV
94266940 WQLHCRSLTRLTKVYGFWRYSISQQAKTRLNQYSEVLYHLYFKEQAANLL
94264333 WQLHCRSLTRLTKVYGFWRYSISQQAKTRLNQYSEVLYHLYFKERAANLL
297569306 WQRHCRELTRLTKVYGFWRYSTSQQAKMRLNQYSEVLYHLFYKARAEQIR
254416162 WKIHTSRLLSLARIYGFWNYTS-KEKREDLLRYIESLFYLIYKPRAQQLLEEHLKR
332712456 WKIHTTRLLSLARIYGFWNYTS-KANQEDMLRYLEALFHLIYKPRAKKLLEEHLTLPARA
17232477 WKIHTTKLLTLARIYGFWNFTS-KEKREDLLRYLESLFYLIYKPRAQQLLEQHKYR
75908500 WKIHTTKLLTLARIYGFWNFAS-KEKREDLLRYLESLFYLIYKPRAQQLLEQHKYR
186682280 WKIHTSKLLSLARIYGFWNFTS-KENREDLLRYLEALFYLIYKPRAQQLLEQHKYR
119512682 WKIHTTKLLSLARIYGFWNFTS-KENREDLLRYLEALFYLIYKPKAQELLEQHQHR
334117431 WKIHTTKLLTLARTYGFWNYSS-KENREDMLRYIESLFYLIYKPKAKALLAEHANR
218440696 WKIHTSKLLSLSRIYGFWNFIS-KENREDILRYVESLFYLLFKPRAKELLEQHMHR
159031025 WKIHTTRLLSLARIYGFWNHTS-QENREELLRYIETLFYLLFKPRAQHLLEIHGQR
220907171 WKIHTTRLLSLARTYGFWNYSL-QENREDLLRYIESLFYLLFKPRAQQLLAEHLQR
22298591 WKIHCTRLLSLAKIYGFWNFSS-QENREDMMRYMEALFHLLYKPRAQALLAEHLQR
158339122 WKIHTSRLLSLAKIYGFWNYTS-REKREDMLRYIETIFYLLYKPMAKKLLAKHMDS
284053161 WKIHTSRLLTLTRVYGFWKYVS-KEKRADMMRYLEALFYLIYKPRCQELLQQHLQR

```
209523126          WKIHTSRLLTLTRVYGFWKYIS-KEKRADMMRYLEALFYLIYKPRSQELLQQHLQR----------------------------------------
17228554           WQLHTSQLLALTKIYSFWNFIR-PESSEARVRYMESLFHLIYKPRAEQILAKHMSCH----------------------------------------
75909957           WQLHTSQLLALTKIYSFWNFIR-PESSEARVRYMESLFHLIYKPRAEQILAKHMSCH----------------------------------------
186685043          WHLYSNQLLLLAKMFTFWNFVA-PENNEARDRYMETLFHLIYKPRAEKILEKHMGANS----------------------------------------
119510624          WHLHTSQILLLAKIFSFWNFAL-PENNAAKHRYLETLFYLIFKPRAEKILEKHQQYSVIMNH------------------------------------
298489784          WESYTEQLLLIAKIFSFWNFIV-PEANDARDRYMEILFHLIYKPRAEQIIGQHQQRYAHIYK------------------------------------
220909283          WPSHIRQLLLLTKIYGFWNCMA-QQQREALLNYMDALYHLIYKPRAAEILAQHLYR----------------------------------------
332708740          WQLHTKQLLSLAKIYGFWSETS-KESREALLRYLEALFYLIYKPRATNLLAKHRSR----------------------------------------
Clustal Consensus                 :

                          910       920       930       940       950       960       970       980
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
114331077         ----------------------------------------------------------------------------------------------------
30249199          ----------------------------------------------------------------------------------------------------
82703384          ----------------------------------------------------------------------------------------------------
198283392         ----------------------------------------------------------------------------------------------------
344199811         ----------------------------------------------------------------------------------------------------
255021595         ----------------------------------------------------------------------------------------------------
77166514          ----------------------------------------------------------------------------------------------------
300115586         ----------------------------------------------------------------------------------------------------
292493898         ----------------------------------------------------------------------------------------------------
350552412         ----------------------------------------------------------------------------------------------------
220933364         ----------------------------------------------------------------------------------------------------
350574485         ----------------------------------------------------------------------------------------------------
95929190          ----------------------------------------------------------------------------------------------------
298528445         ----------------------------------------------------------------------------------------------------
297602308         ----------------------------------------------------------------------------------------------------
115457664         E---------------------------------------------------------------------------------------------------
326504012         EQNGYR----------------------------------------------------------------------------------------------
115450038         ----------------------------------------------------------------------------------------------------
326531526         NFSQDAA---------------------------------------------------------------------------------------------
242063616         ----------------------------------------------------------------------------------------------------
42568160          ----------------------------------------------------------------------------------------------------
297805240         ----------------------------------------------------------------------------------------------------
313770765         ----------------------------------------------------------------------------------------------------
313770767         ----------------------------------------------------------------------------------------------------
255584097         ----------------------------------------------------------------------------------------------------
225466221         ----------------------------------------------------------------------------------------------------
313770769         ----------------------------------------------------------------------------------------------------
313770771         ----------------------------------------------------------------------------------------------------
255570671         KPETITTRHVI-----------------------------------------------------------------------------------------
15219457          HDVKVGEREVREGLLAADASERVKKVLESSEEKQKLEKMKIAYGQQHSQGASPVRNLFWSVVVCLYICYILKQRFFGANSAQEY
297842089         NDGKVGERDVSETFLAAEASERTKKVLETSEEKQKLEKMKIAYGQQQNQGASPVRNLFWSVVVCLYICYILKQRFFGANSAQEY
115466896         ----------------------------------------------------------------------------------------------------
17980243          ----------------------------------------------------------------------------------------------------
162460741         ----------------------------------------------------------------------------------------------------
239984690         ----------------------------------------------------------------------------------------------------
```

162460681	----------
242035533	----------
62865493	----------
17980241	----------
115453437	----------
115473359	----------
3980298	----------
15242073	----------
297812265	----------
22331535	----------
297818772	----------
313770759	----------
313770761	----------
255550319	----------
324984223	----------
258489633	----------
324984221	----------
324984229	----------
6683114	----------
4584690	----------
3915046	----------
3766299	----------
3377764	----------
225444613	----------
3660531	----------
31455440	----------
304651488	----------
28629438	----------
115310618	----------
345286417	----------
38425095	----------
15235300	----------
297814081	----------
6682841	----------
255564236	----------
313770763	----------
3915037	----------
225437428	----------
162458268	----------
242035817	----------
115452927	----------
241896730	----------
304651490	----------
29289943	----------
115310620	----------
4468153	----------
345286419	----------

```
4468151          ------------------------------------------------------------------------------
151176306        ------------------------------------------------------------------------------
14530225         ------------------------------------------------------------------------------
15239816         ------------------------------------------------------------------------------
297795665        ------------------------------------------------------------------------------
225431790        ------------------------------------------------------------------------------
255551835        ------------------------------------------------------------------------------
168009716        EGPEEKTETKARLGPGQAAIVGTPASA----------------------------------------------------
168035060        ------------------------------------------------------------------------------
168058907        ------------------------------------------------------------------------------
168029793        ------------------------------------------------------------------------------
94266940         ------------------------------------------------------------------------------
94264333         ------------------------------------------------------------------------------
297569306        ------------------------------------------------------------------------------
254416162        ------------------------------------------------------------------------------
332712456        ------------------------------------------------------------------------------
17232477         ------------------------------------------------------------------------------
75908500         ------------------------------------------------------------------------------
186682280        ------------------------------------------------------------------------------
119512682        ------------------------------------------------------------------------------
334117431        ------------------------------------------------------------------------------
218440696        ------------------------------------------------------------------------------
159031025        ------------------------------------------------------------------------------
220907171        ------------------------------------------------------------------------------
22298591         ------------------------------------------------------------------------------
158339122        ------------------------------------------------------------------------------
284053161        ------------------------------------------------------------------------------
209523126        ------------------------------------------------------------------------------
17228554         ------------------------------------------------------------------------------
75909957         ------------------------------------------------------------------------------
186685043        ------------------------------------------------------------------------------
119510624        ------------------------------------------------------------------------------
298489784        ------------------------------------------------------------------------------
220909283        ------------------------------------------------------------------------------
332708740        ------------------------------------------------------------------------------
Clustal Consensus
```

SucSase sequences from *P. patens* subsp. patens (group IV) are clearly separated from those of vascular plants (Figure 35). Interestingly, groups V and VII are further divided in two major branches, containing the sequences from dicots (green) and monocots (blue), respectively. This separation is less clear in group VI (Figure 35). SucSase from *N. europaea* is in a small branch with other β-proteobacteria in group III (proteobacteria). Clearly, *Neu*SucSase is well separated from plant and cyanobacterial enzymes, though they share a significant similarity among all of them. For instance the identity between *Neu*SucSase and *Ath*SucSase is 52.2% and between *Neu*SucSase and SucSase from *T. elongatus* is 48.5%.

Protein Expression and Characterization

The gene of the putative SucSase in *N. europaea* (NCBI Protein ID NP_841269) codes for 794 amino acids. To shed light on Suc metabolism of group III (Figure 35), we amplified this sequence and expressed the recombinant protein in *E. coli* cells. The enzyme was purified to homogeneity by HisTrap column and gel filtration chromatography as mentioned in "Materials and Methods". The recombinant protein migrated in SDS-PAGE as a single band of ~95 kDa (data not shown), which is in good agreement with the predicted molecular mass of 93 kDa (including the His-tag provided by the pET28c vector). *Neu*SucSase eluted from the Superdex 200 (size exclusion) column as a protein of ~360 kDa (data not shown), suggesting a tetrameric quaternary structure, as it was reported for cyanobacterial and plant SucSases [77, 80, 84].

\

Substrate Specificity of *Neu*SucSase

SucSases from plants have shown a certain degree of promiscuity to transfer glucoses from ADP-Glc and UDP-Glc, though UDP-Glc is generally preferred. We tested the substrate specificity of SucSase from *N. europaea* in the Suc synthesis direction (Table 4), and observed that ADP-Glc is a more efficient substrate than UDP-Glc. The main difference is not given by V_{max}, but by a higher apparent affinity towards ADP-Glc. The $S_{0.5}$ for ADP-Glc is 0.044 mM in presence of optimal concentrations of Fru (20 mM); whereas the $S_{0.5}$ for UDP-Glc is 0.98 mM in presence of optimal concentrations of Fru (500 mM). On the other hand, the apparent affinity for Fru is higher in presence of ADP-Glc rather than UDP-Glc. The $S_{0.5}$ for Fru at saturated concentrations of ADP-Glc is 5.6 mM whereas the $S_{0.5}$ for Fru in presence of UDP-Glc is significantly higher. Because of the high concentrations of Fru needed to reach saturation, it is not possible to measure the $S_{0.5}$ for Fru with high precision; but it is at least ~20-fold higher (120 mM). The catalytic efficiencies calculated for ADP-Glc and $Fru_{(ADP\text{-}Glc)}$ were 17- and 37-fold higher than those obtained for UDP-Glc and $Fru_{(UDP\text{-}Glc)}$, respectively (Table 4). These results indicate that *Neu*SucSase prefers ADP-Glc over UDP-Glc as substrate. Similar conclusions were obtained for the enzyme from the cyanobacterium *T. elongatus*, which showed a 26-fold higher catalytic efficiency for ADP-Glc than UDP-Glc [77].

Table 4. Kinetic parameters for substrates of *Neu*SucSase in the direction of sucrose synthesis. Assays were performed using the conditions described in the "Materials and Methods" section. Analogous values to catalytic efficiency ($V_{max}/S_{0.5}$) were calculated using the predicted molecular mass of 93 kDa.

Substrate	$S_{0.5}$ (mM)	V_{max} (U mg^{-1})	n_H	$V_{max}/S_{0.5}$ (mM^{-1} s^{-1})
UDP-Glc	0.89 ± 0.05	4.3 ± 0.1	1.1	7.5
ADP-Glc	0.044 ± 0.006	3.7 ± 0.1	1.3	130.3
$Fru_{(UDP\text{-}Glc)}$	120 ± 10	2.8 ± 0.2	1.3	0.036
$Fru_{(ADP\text{-}Glc)}$	5.6 ± 0.4	4.8 ± 0.2	1.6	1.33

X-ray Diffraction, Data Processing, Model Building, and Refinement

The best data set collected at synchrotron beamline was processed to 3.05 Å and indexed as space group P65. It was integrated and scaled producing good statistics (Table 5). After the molecular replacement search, four copies of the starting model described in "Materials and Methods" were found in one asymmetric unit. Iterative cycles of model building and refinement were conducted yielding a well-defined structure for *Neu*SucSase with R_{work} and R_{free} values of 17.37 % and 21.75 %, respectively (Table 5). The truncated GT-B(D) domain was built according to the electron density map. The final structural model contains all the residues except the first three at the N-terminus and the last two at the C-terminus of the *Neu*SucSase amino acid sequence (Figure 36).

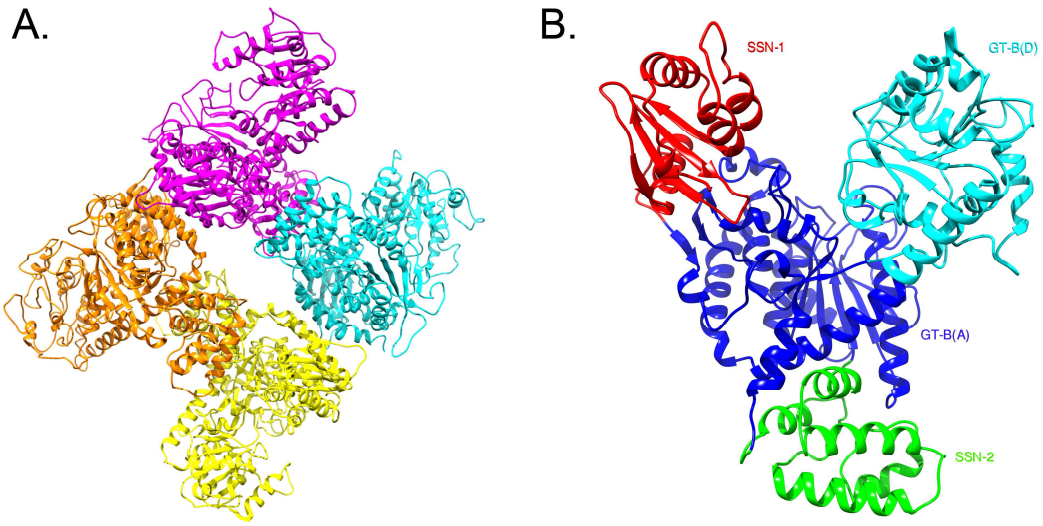


Figure 37. The crystal structure of *Neu*SucSase. A. Tetrameric structure of *Neu*SucSase. B. Monomeric structure and its four different domains: SSN-1, SSN-2, GT-B(A), and GT-B(D).

Structural Analysis of *Neu*SucSase

Overall structure. Although the resolution of the data set was 3.05 Å, the backbone of the protein and some of the key residues side chains were well defined by the electron density (Figure 38 and Figure 39). This allowed us to conduct detailed structural analysis on SucSase's conformational changes involving backbone movement, which are relevant to the catalytic cycle. The crystal structure of *Neu*SucSase displayed a similar fold to the previously reported structural model of *Ath*SucSase (PDB ID 3S29) [84]. *Neu*SucSase is a tetramer composed of four identical subunits (Figure 36A), where each monomer contains four domains (Figure 36B). The first domain designated as "Sucrose Synthase N-terminal-1" (SSN-1) domain included residues 1-119 (Figure 36B, red) and contained five α-helixes and four β-strands. The second domain, which included residues 159-264, is the "Sucrose Synthase N-terminal- (SSN-2) domain (Figure 36B, green) with five α-

helixes. The third is the GT-B(D) domain (Figure 36B, cyan), which included residues 514-742 with eight α-helixes and three β-strands. This domain functions in the GT-B family as the sugar donor (D) in catalysis [84, 86, 91]. The fourth domain is the GT-B(A) domain (Figure 36B, blue), which consists of residues from three separate regions: 120-158, 265-513, and 743-794. These separate regions are encompassed by the SSN-1, SSN-2 and GT-B(D) domains in the center of the monomer. The GT-B(A) domain included nine α-helixes and eight β-strands and functions in the GT-B family as the sugar acceptor (A) in catalysis [84, 86, 91]. Both the GT-B(D) and GT-B(A) domains directly facilitate the glycosyltransferase reaction [84].

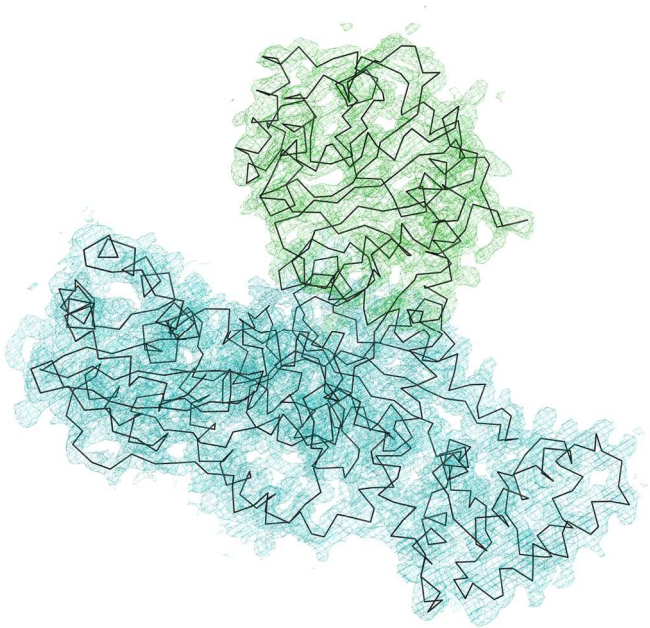


Figure 38. The *Neu*SucSase monomer structure is shown in backbone trace. The electron density (2Fo-Fc map at 1 sigma) is shown as mashes. The green portion of the electron density covers the GT-B (D) domain in the "open" position, while the cyan portion of the electron density covers the SSN-1, SSN-2, and GT-B (A) domains.

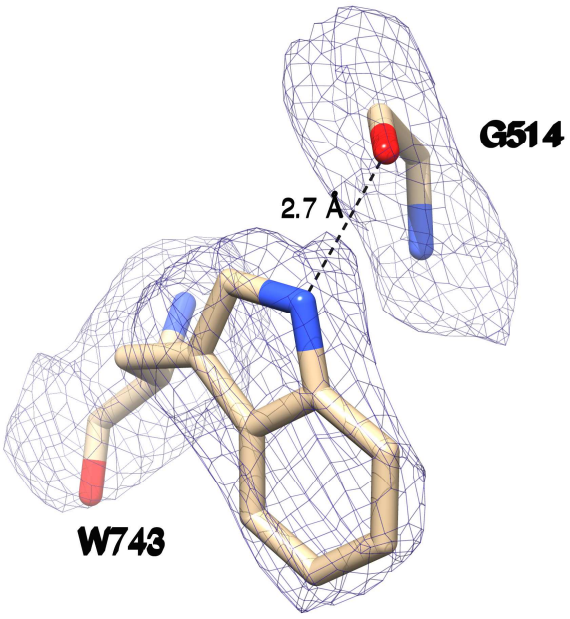


Figure 39. The "hinge". The picture displays the hydrogen bond between W73 and G514. The electron density (2Fo-Fc map) is shown as blue mesh at 1.5 sigma level.

The obtained *Neu*SucSase structure with no substrates bound has a clearly different overall conformation when compared to *Ath*SucSase with UDP and Fru [84]. This implies that substrate binding induces significant conformational changes (Figure 40). After superimposition of only the GT-B(A) domains of *Ath*SucSase and *Neu*SucSase (using the least squares function in COOT), the SSN-1, SSN-2, and GT-B(A) domains from both structures overlapped well while the GT-B(D) domains were in a different relative position. The angle between the GT-B(A) and GT-B(D) domains in *Neu*SucSase was about 23.5 degrees wider than in the *Ath*SucSase structure. Based on such comparison, we suggest that the *Neu*SucSase structure in this work was in an "open" conformation whereas the *Ath*SucSase was "closed" [84]. We have identified some distinct

structural determinants (hinges and latches) related to the movements of the sugar (GT-B(A)) and nucleotide (GT-B(D)) binding domains.

Sugar-binding GT-B(A) domain. Analysis of a difference distant matrix map of the Fru-binding GT-B(A) domain as described in "Materials and Methods" (Figure 41) highlights three main regions that move closer upon sugar binding. These are 325-375 to 280-290 (~5 Å), 425-435 to 280-290 (~4 Å), and 425-435 to 325-375 (~3 Å) (Figure 41). Other pair of regions that move towards each other are 280-290 to 490-505 (~3 Å) and 280-290 to 450-460 (~2 Å) (Figure 41). From this analysis, the area 280-290 is the most involved in an induced fit interaction with Fru. Further inspection of these areas reveals that Fru induces local conformational changes via superimposition of the GT-B(A) domains of the *Ath*SucSase (closed) and the *Neu*SucSase (open) (Figure 42). These include the side chain of K431 and the backbone of residues 288-290. The re-shaping of the Fru binding site facilitates the closing via a set of inter-domain hydrogen bonds (Figure 42 in green). These local conformational changes along with the presence of Fru further promote the interactions between the GT-B(A) and GT-(D) domains. Thus, we

propose that Fru binding contributes to the induced fit closing of the structure.

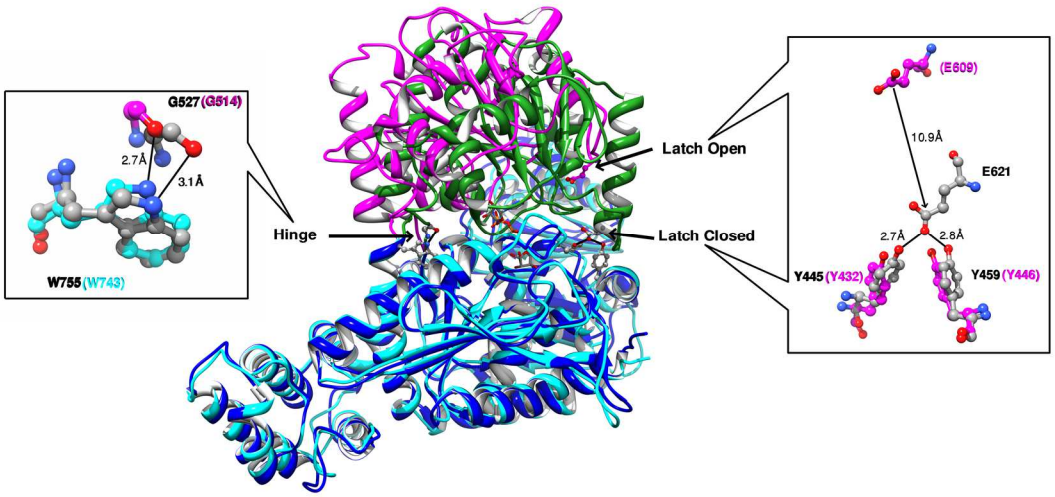


Figure 40. Comparison of the open and closed monomeric forms. The open form structure is represented by the *Neu*SucSase structure reported in this paper; the closed form is represented by the *Ath*SucSase structure (PDB ID 3S29). The SNN-1, SNN-2 and GT-B(A) domains are shown in blue for the open form structure and in cyan for the closed form structure. The GT-B(D) domain is shown in magenta for the open form and in green for the closed form. The "hinge-latch" features of the domain movement are shown in blown-up views.

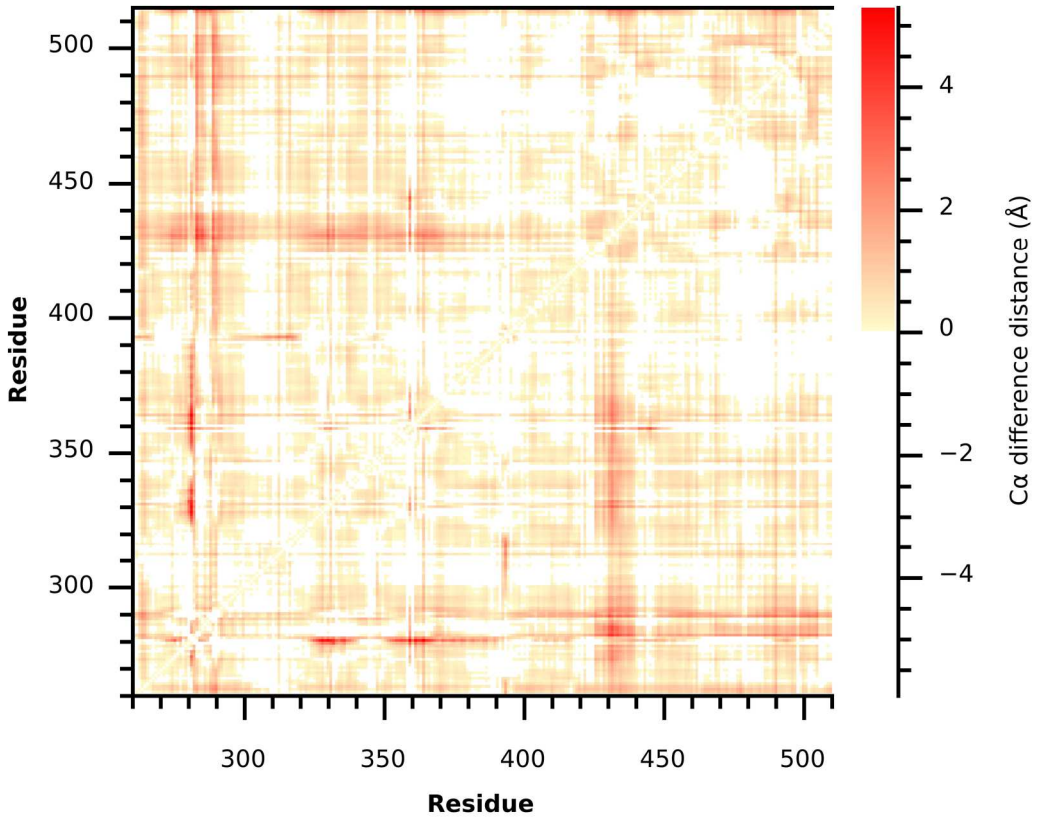


Figure 41. Difference distance matrix map of the GT-B(A) domain. Distances were calculated between all pair of Cα carbon of the open structure (*Neu*SucSase). A second pairwise distance matrix was calculated for the close structure (homology model of *Neu*SucSase using *Ath*SucSase as template as described in "Materials and Methods"). Afterwards, these two matrices were subtracted, and the Δdistance was color coded. The negative and zero values are represented in white. Red colors (higher Δdistance values) are pairs of Cα carbon that are getting closer upon closing of the enzyme. Only residues from 260 to 510 are shown, which correspond to the GT-B(A) domain.

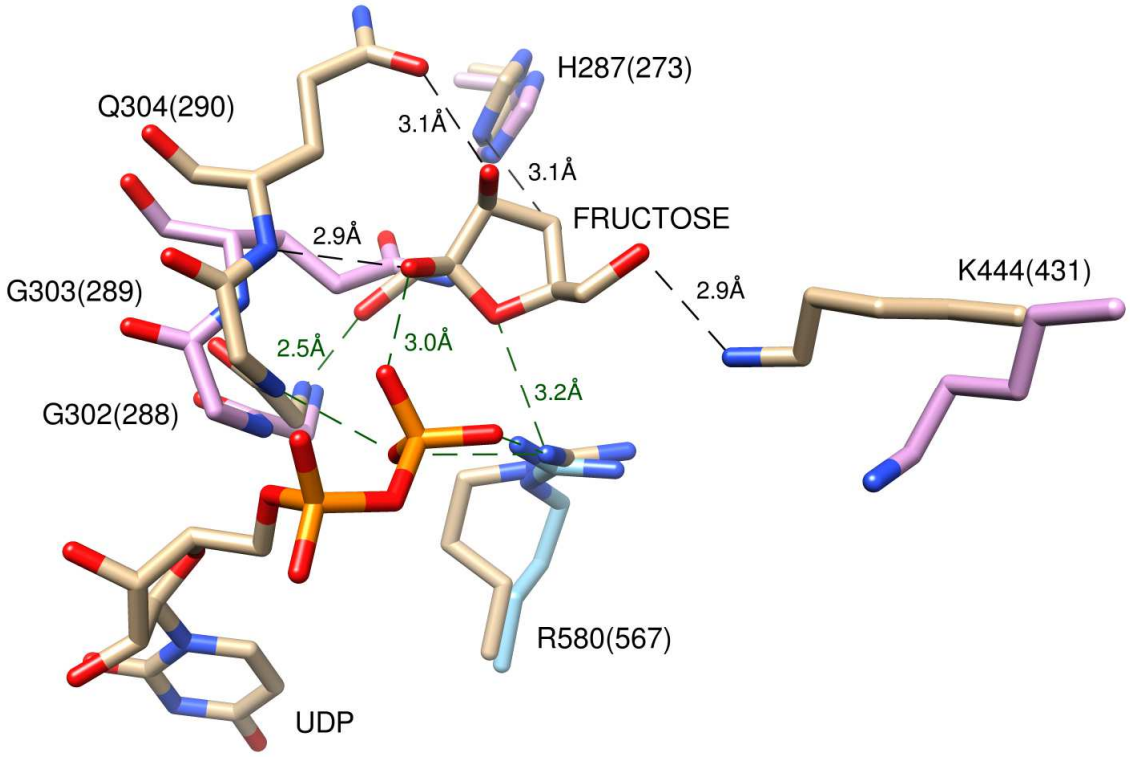


Figure 42. Overlap comparison of the fructose binding sites of the open (*Neu*SucSase) and closed (*Ath*SucSase, PDB ID 3S29) structures. The carbon atoms in the closed form structure are in pale yellow. The carbon atoms in the open form structure are in cyan (GT-B(A) domain) and pink (GT-B(D) domain). Conserved residues between two structures are labeled with respective residue numbers; the residue numbers of the open form structure (*Neu*SucSase) are in parenthesis. The hydrogen bonds in the GT-B(D) domain are shown in black. The hydrogen bonds in between GT-B(A) and GT-B(D) domains and the ones in GT-B(A) domains are shown in green.

Nucleotide-binding GT-B(D) domain. The GT-B(D) domain binds to sugar nucleotide (synthesis direction) or nucleotide (cleavage direction) substrates. When overlapping GT-B(D) domains from both *Ath*SucSase and *Neu*SucSase, residues interacting with the phosphate and ribose moieties of the nucleotides are not only conserved, but also at the same positions (Figure 43). On the other hand, two residues interacting with the nucleotide base are not conserved. Residues Q648 and N654 from *Ath*SucSase are replaced by R636 and A642 in *Neu*SucSase, respectively. This difference

creates a more spacious binding site in *Neu*SucSase, which may accommodate bulkier adenosine nucleotide substrates. Similar sequence differences were observed in the SucSase from *T. elongatus* [77]. Based on sequence analysis and homology modeling it was suggested that these two residues could be responsible for the preference towards ADP/ADP-Glc over other nucleotides such as UDP/UDP-Glc in the cyanobacterial enzyme. It is important to notice that the side chains in R636 and A642 in *Neu*SucSase are not conserved in the *E. coli* Glycogen synthase, which is another glycosyltransferase that binds ADP-Glc [92]. *E. coli* Glycogen synthase has a different motif in that position with a Tyr and Ser instead of Arg and Ala [86, 92, 93], implying a different structural arrangement for accommodating ADP-Glc. Overall, the nucleotide binding to the GT-B(D) domain does not seem to trigger significant local conformational changes (Figure 43). The direct interactions with the nucleotide do not make major contributions to the induced fit mechanism.

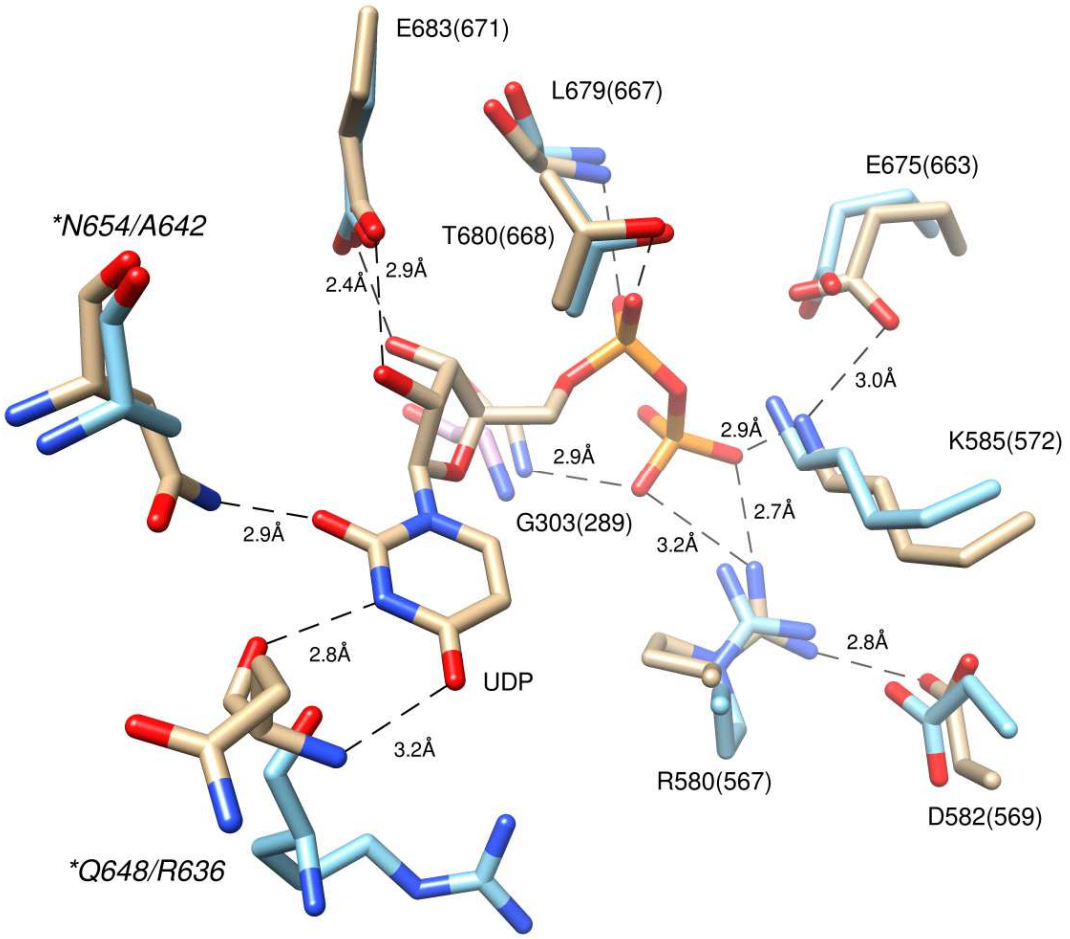


Figure 43. Overlap comparison of the nucleotide binding sites of the open (*Neu*SucSase) and closed (*Ath*SucSase, PDB ID 3S29) structures. The carbon atoms in the closed form structure are in pale yellow. The carbon atoms in the open form structure are in cyan (GT-B(A) domain) and pink (GT-B(D) domain). Conserved residues between two structures are labeled with respective residue numbers; the residue numbers of the open form structure (*Neu*SucSase) are in parenthesis. The asterisks indicate the non-conserved binding residues with the closed form residues labeled in front of the ones in the open form structure.

Hinge Analysis. We scanned the structures of the acceptor and donor domains for hinges and subtle conformational changes that could be functionally important in catalysis. We used the in-house program *hingescan* described in "Materials and Methods". Using several window sizes we detected several local conformational changes (Figure 44). For a window size of 51, we detected two clear hinge elements near residues

~515 and ~744 (Figure 45). These two elements actually form a single "hinge" that comprises a hydrogen bond between two conserved residues (G514 and W743) in a flexible area (Figure 40, Figure 39, and Figure 44). These two residues remain at the same position in both the open and the closed conformations of the enzyme. For smaller windows, we found other significant local secondary structure rearrangements (Figure 44, Figure 46, and Figure 45). Upon closing, two α-helices in the GT-B(D) domain are extended, and a β-strand replaces a previously coiled stretch. The outcome is a more ordered peripheric structure of the GT-B(D) domain. We propose that this secondary structure rearrangement, despite the local entropy decrease, would release extra energy to close the conformation facilitating the binding of substrates.

Latches. A feature that contributes to the stabilization of the closed form is a "latch", E609, which comprises the highly conserved E609 residue located at the periphery of the GT-B(D) domain (Figure 37). Going from an open to a closed conformation, this glutamate residue moves ~11 Å towards the GT-B(A) domain and ends up hydrogen-bonded to two tyrosine residues (Y432 and Y446) stabilizing the closing (Figure 40 and Figure 47). Interestingly, there were small secondary structure

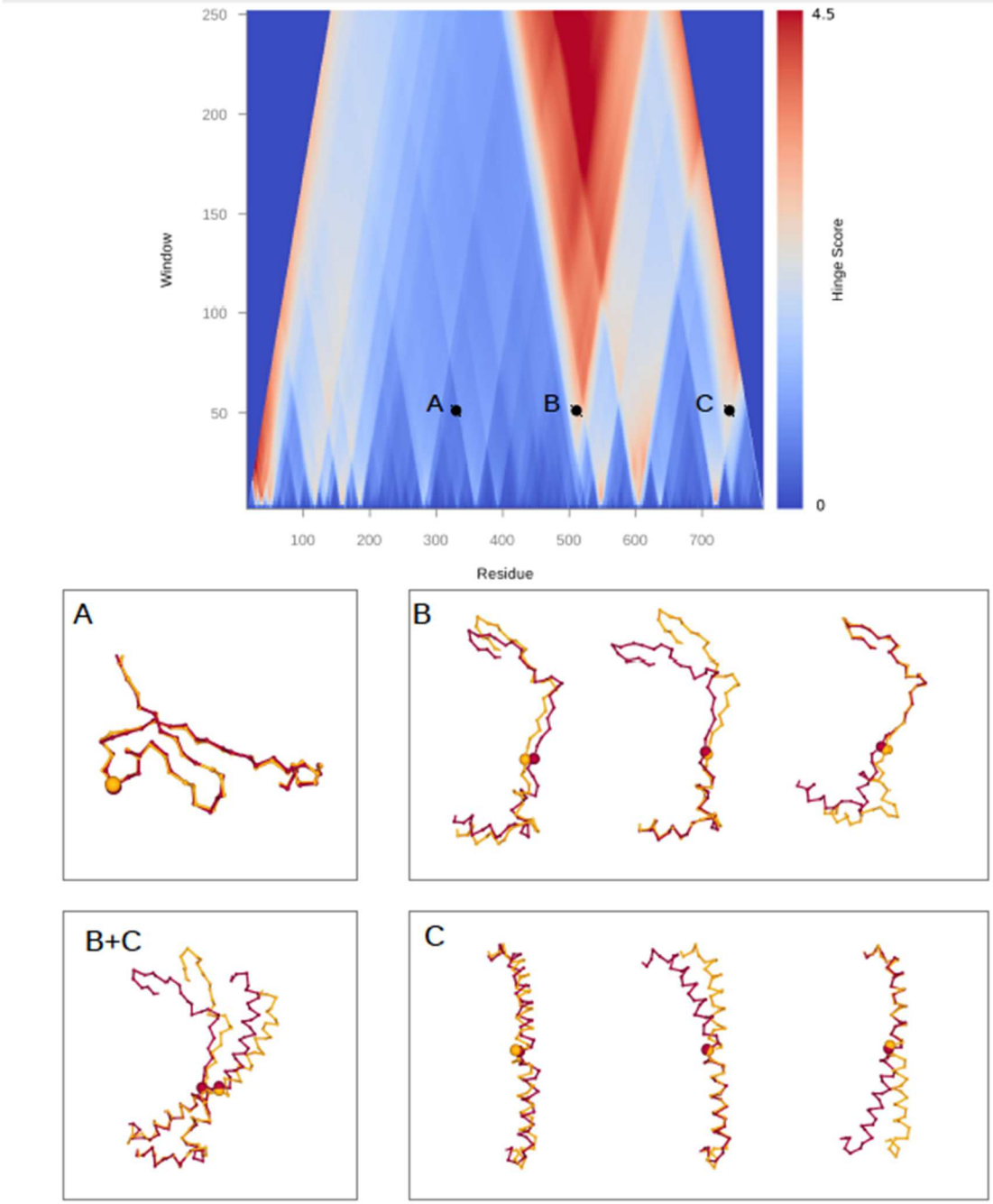


Figure 44. Hinge analysis of *Neu*SucSase . Comparison between the open form and the modeled closed form of the enzyme was performed with the program hingescan as described in Materials and Methods. In the top panel, a color coded hinge score (0 to 4.5 .) is plot for different window sizes at each residue. Three points are highlighted, A (control) and two hinge regions that are near in space (B and C). In all four panels (A, B, C, and B+C), maroon structures belong to the closed form, and yellow ones to the open form. Panel A shows optimal rigid body superposition of regions 307-357 of the open and closed form. Panel B displays regions 489-539 of both the close and open form.

The left, center and right overlaps are optimal rigid body superposition of the Cα of residues 489-539, 514-539, and 489-514, respectively. The center and right overlaps were performed to highlight the differences on the other half of the region. Spheres represent center residue 514.Panel C displays regions 718-768 of both the close and open form. The left, center and right overlaps are optimal rigid body superposition of the Cα of residues 718-768, 718-743, and 743-768, respectively. The center and right overlaps were performed to highlight the differences on the other half of the region. Spheres represent center residue 743. Panel B+C shows both regions (489-539 and 718-768), in which the optimal superposition was performed based on Cα of residues 514-539 and 718-743.

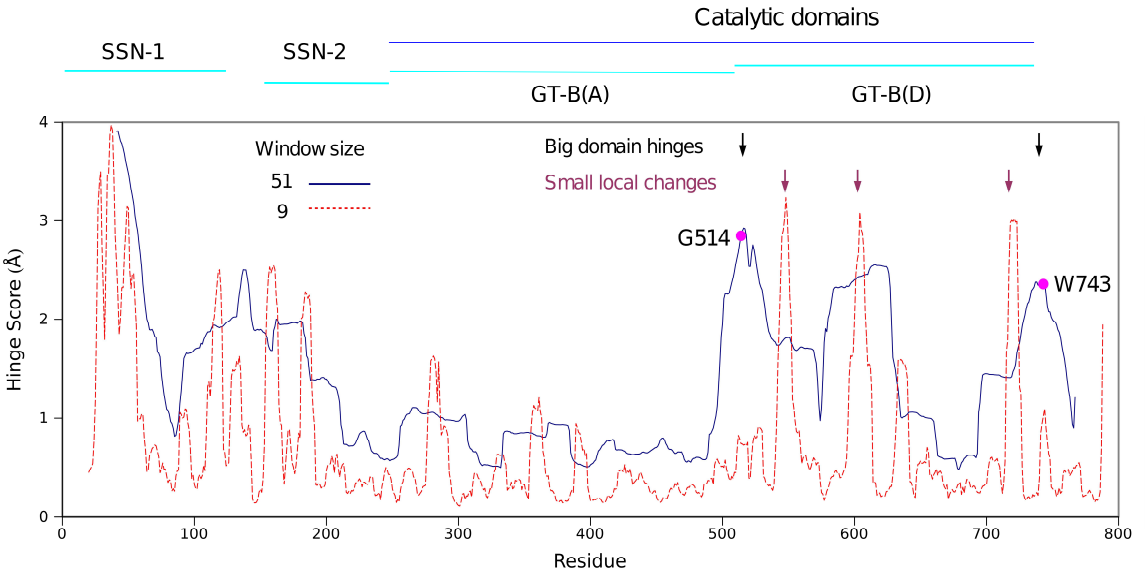


Figure 45. Hinge analysis by comparison of the open v. closed conformations. The blue and red show the "hinge score" using 51 and 9 windows, respectively. The magmata dots (also point with black arrows) shows the two distinct hinges: G514 and W743. The purple arrows points at the region displaying the secondary structure rearrangements.

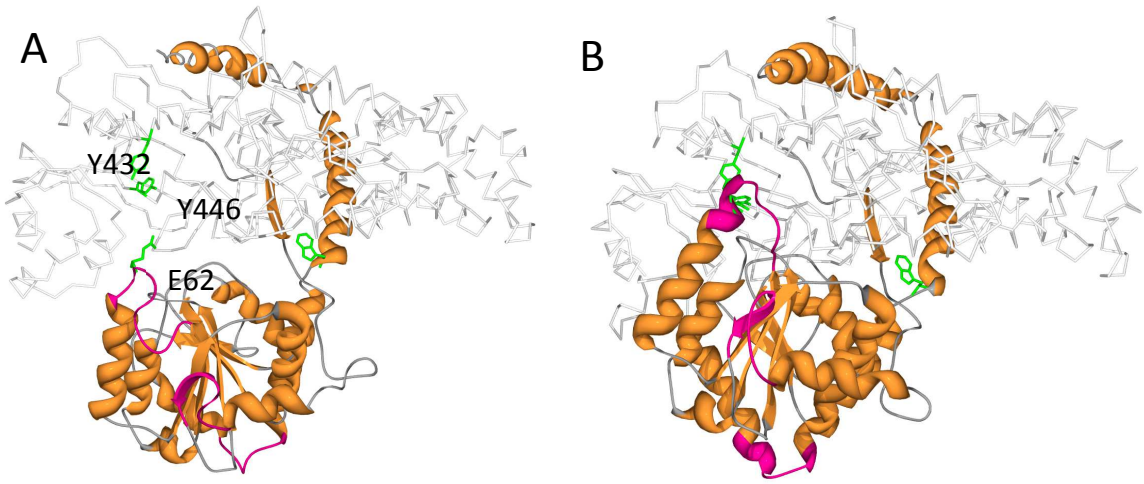


Figure 46. Secondary structure rearrangement during conformational changes between open and closed structures of bacterial SucSases. A. Open conformation of the *Neu*SucSase structure. B. Closed conformation of the *Ath*SucSase structure (PDB ID 3S29). The GT-B(D) domain and a long bending helix is show in orange. The SNN-1, SNN-2, and GT-B(A) domains are in grey. The side chains of the residues in the "latch" and the "hinge" are shown as green sticks. The secondary structural rearrangement is shown in magenta. T

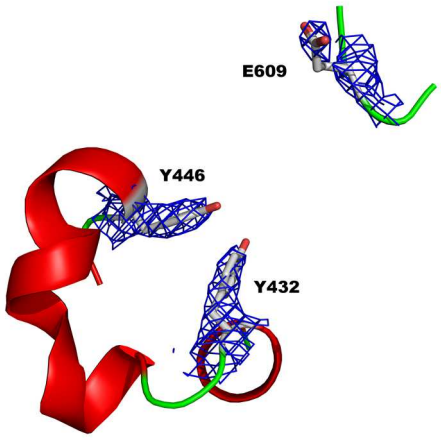


Figure 47. The open "latch". E609, Y446 and Y432 are shown in sticks. The electron density (2Fo-Fc map) is shown as blue mesh at 1.0 sigma level around there key residues. Local secondary structure fetures are shown in cartoon style.

rearrangements in the vicinity of this latch, which could facilitate the interaction between E609 and the two tyrosines (Figure 46). On the other side of the active site, opposite to

the latch described, there is a hydrophobic patch that contributes to seal the closing (Figure 48). Two hydrophobic residues (M635, L637) in the GT-B(D) domain get in contact with a hydrophobic cluster (V281, L282 and L284) in the GT-B(A) domain, upon closing. The side chain of N280 also provides a methylene to build a non-polar pocket that latches on to M635 and L637 (Figure 48). On the other hand, the amide polar group is exposed to the solvent.

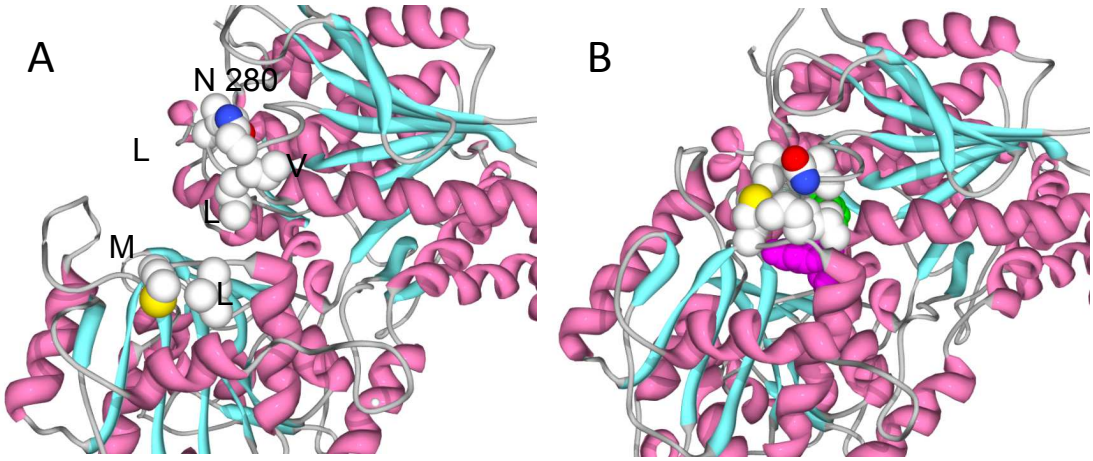


Figure 48. Hydrophobic residues contribute to the latch action. Panel A depicts the open (*Neu*SucSase crystal structure) and B a homology model of a closed structure, which was built as described in the "Materials and Methods" section. Upon closing, the hydrophobic residues: M635, L637 in the GT-B (D) domain and N280, V281, L282 and L284 in GT-B(A) domain generate a hydrophobic environment that stabilize the close action.

The closed structure seems to induce stronger interactions with the nucleotide and vice versa. In *Neu*SucSase, the conserved E671 is in the same position as E369 in the *E. coli* trehalose-6-phosphate synthase (OtsA) [92, 94]. In OtsA, as well as in the close conformation of *Ath*SucSase (E683), the carboxylate of this side chain forms two hydrogen bonds with the hydroxyl groups of the ribose of the nucleotide. In these enzymes, the carboxylate is surrounded by hydrophobic residues (Y520, Y646, L667,

T668 in *Neu*SucSase, Y533, Y658, L679 and T680 in *Ath*SucSase), which makes the hydrogen bonds stronger in the non-polar environment. In the open form of the *Neu*SucSase structure, V291 (V306 in *Ath*SucSase) moves away from the side chain of the glutamate residue (Figure 49). This implies that the closing recruits a non-polar side chain to completely surround the carboxylate. Consequently, nucleotide binding stabilizes he carboxylate charge and facilitates the interaction with V291 upon closing. Interestingly, in another glycosyltransferase such as bacterial glycogen synthase, E671 has been replaced by a Tyr, and V291 was replaced by Asp, thus, switching their roles [92]. Therefore, this ligand-dependent interaction may be a common feature in this family of enzymes.

Site Directed Mutagenesis of Critical Residues

Previously, important residues for catalysis were identified in other retaining GT-B glycosyltransferases. A triad of critical residues has been found in the active site of maltodextrin phosphorylase [95, 96] and glycogen synthase from *E. coli* [97]. Based on X-ray structures, these residues were also predicted to be important for catalysis in OtsA [94] and *Ath*SucSase [84]. The homologue residues in *Neu*SucSase are R567, K572, and E663 (Figure 50). When any of those residues were replaced by alanine the activity in the direction of Suc synthesis severely decreased, either in presence of ADP-Glc or UDP-Glc (Table 6). The most active of all these mutants was E663A in presence of ADP-Glc, but it was still 200-fold less active than the wild type. These results indicate that this triad is critical in *Neu*SucSase. Consequently, it also suggests that *Neu*SucSase together with all

other retaining glycosyltransferases with a GT-B fold share the same reaction mechanism (Figure 50).

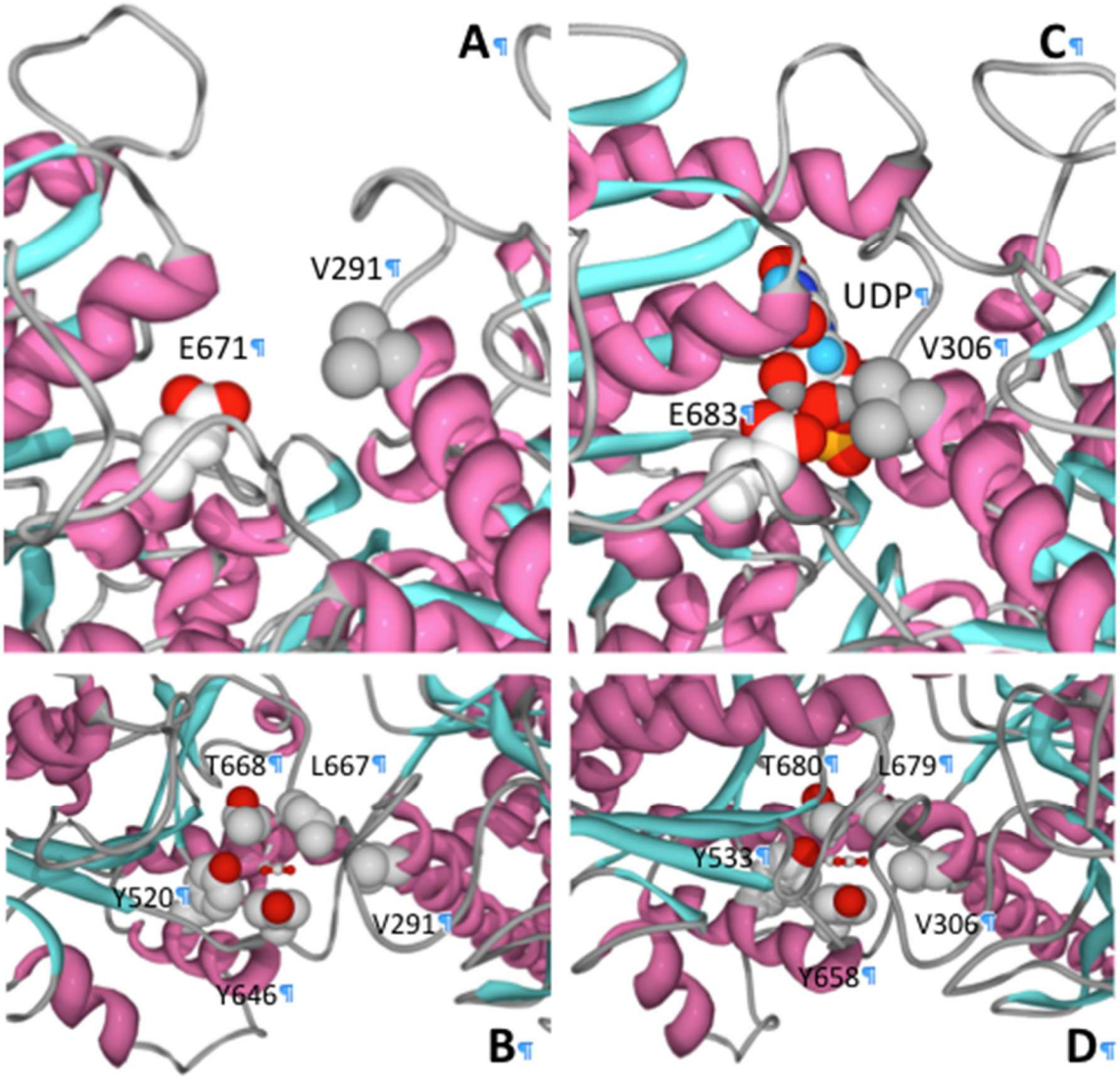


Figure 49. Conserved hydrophobic residues are involved in substrate binding in both open conformation of *Neu*SucSase (A and B), and closed conformation of *Ath*SucSase (PDB ID: 3S29) (C and D). In the open conformation, V291, E691 two conserved hydrophobic residues are shown in sticks in figure A, and V291, Y520, Y646, L667, T668 (rotate 90 degree from A) are shown in sticks in figure B. In the closed conformation, V 306 and E 683 are conserved hydrophobic residues in figure C, and V306, Y533, Y658, L679 and T680 (rotate 90 degree from C) are shown in sticks in figure D.

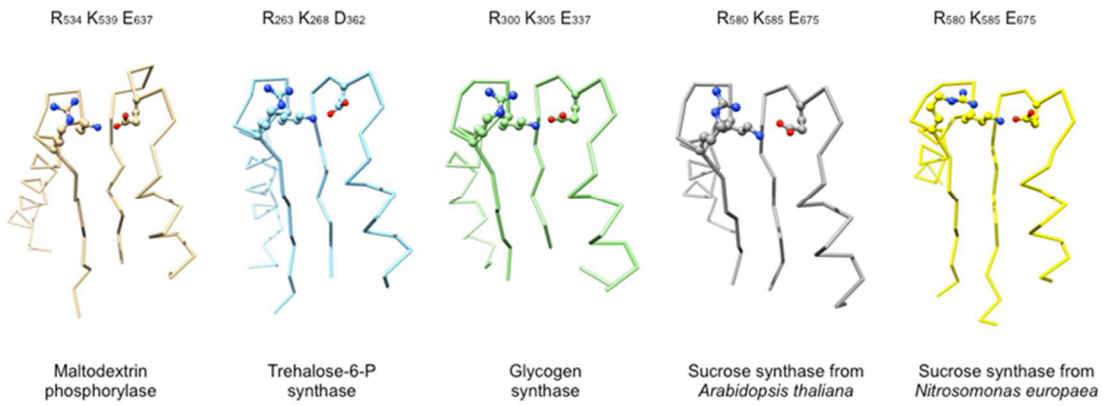


Figure 50. Three highly conserved catalytic residues in different members of the retaining GT-B glycosyltransferase family. The structures analysed in this figure are maltodextrin phosphorylase (PDB ID 1E4O), trehalose-6-phosphate synthase (PDB ID 1GZ5), and glycogen synthase (PDB ID 2ZQS) from *E. coli*, *Ath*SucSase (PDB ID 3S29), and *Neu*SucSase (this work).

Table 5. Data collection and refinement statistics for *Neu*SucSase

Data Processing	
Space group	P65
Cell dimensions	
a; b; c (Å)	236.90; 236.90; 231.44
α; β; γ (°)	90.00; 90.00; 120.00
Resolution (Å)	3.05
Mosaicity (°)	0.47
[a]R_{merge}	0.169 (0.963)
I/sigma	9.6 (2.1)
Completeness	97.9 (98.6)
Multiplicity	6.3 (6.3)
Refinement	
Resolution (Å)	3.05
No. reflections	794715
No. unique reflections	126170
[b]R_{work}/[c]R_{free}	17.37/21.75
[d]RMSD Bond length (Å)	0.009
RMSD Bond angle (°)	1.439

The values for the highest resolution bin are in parentheses.
[a]Linear $R_{merge} = \Sigma|I_{obs}-I_{avg}|/ \Sigma I_{avg}$
[b]$R = \Sigma|F_{obs}-F_{calc}|/ \Sigma F_{obs}$.
[c]Five percent of the reflection data were selected at random as a test set and only these data were used to calculate R_{free}.
[d]RMSD, root mean square deviation.

Table 6. Activity of wild type and mutants of *Neu*SucSase. Assays were performed using the conditions described in the "Materials and Methods" section.

Substrate	V_{max} (U mg^{-1})			
	WT	R567A	K572A	E663A
UDP-Glc	4.3 ± 0.1	< 0.0017	< 0.0019	< 0.01
ADP-Glc	3.7 ± 0.1	< 0.0014	< 0.0016	0.020 ± 0.02

Materials and Methods

Materials: Chemicals and coupled enzymes used for activity assays were from Sigma-Aldrich (St. Louis, MO). *Escherichia coli* BL21 (DE3) cells were purchased from New England BioLabs (Ipswich, MA). Bacterial growth media and antibiotics were from Fisher Scientific (Pittsburgh, PA) and Sigma-Aldrich. Crystallization screen solutions and other supplies were purchased from Hampton Research (Aliso Viejo, CA) and Emerald Bio (Bedford, MA). All the other chemicals were of the highest quality available.

Cloning: The sequence coding for *Neu*SucSase was amplified by PCR using genomic DNA from *N. europaea* ATCC 19718 as template, the specific oligonucleotides CATATGACCACGATTGACACACTCGCCACCTGTACCC (forward, *Nde*I site underlined) and GTCGACTCATATCTCATGGGCCAGCCTGTTTGCCAGCGGCC (reverse, *Sal*I site underlined) as primers, and Phusion HF DNA polymerase (Thermo Fisher Scientific, Rockford, IL) following the manufacturer's instructions. The program used included an initial denaturation of 30 s at 98 °C; 30 cycles of 98 °C for 5 s, 50 °C for 20 s, and 72 °C for 2 min; and a final extension of 72 °C for 5 min. The PCR product was purified after agarose gel electrophoresis and inserted into the pSC-B vector using

the StrataClone Blunt PCR cloning kit (Agilent Technologies, Santa Clara, CA). Sequence identity was checked by automated DNA sequencing at CRC (Comprehensive Cancer Center at University of Chicago, IL). Afterwards, the sequence was subcloned into the pET28c vector (Merck KGaA, Darmstadt, Germany) between *Nde*I and *Sal*I sites to obtain pNESS1, which is the plasmid that encodes the recombinant *Neu*SucSase with an N-terminal His_6-tag.

Site-directed mutagenesis: Site-directed mutagenesis was performed by PCR overlap extension as previously described using Phusion DNA polymerase [98, 99]. Plasmid encoding the *Neu*SucSase (pNESS1) was used as a template for mutagenesis.
To introduce mutations in the sequence coding for *Neu*SucSase we used the following primers: TTTACCATGGCGgcgCTGGATCGGATC (forward) and GATCCGATCCAGcgcCGCCATGGTAAA (reverse) for mutant R567A; CTGGATCGGATCgcgAACATTACCGGC (forward) and GCCGGTAATGTTcgcGATCCGATCCAG (reverse) for mutant K572A; and CCAGCCCTGTTCgcgGCATTCGGCCTG (forward) and CAGGCCGAATGCcgcGAACAGGGCTGG (reverse) for mutant E663A. PCR conditions were the same as those described above. Flanking primers for the PCR overlap extension were the same used for cloning (described above). All mutations were confirmed by DNA sequencing.

Protein expression and purification: Transformed *E. coli* BL21 (DE3) cells with pNESS1 were grown in 4 x 1 L of LB supplemented with 100 μg/ml carbenicillin. This was performed in a 2.8 L Fernbach flask at 37 °C and 250 rpm until $OD_{600\,nm}$ reached

~0.6. Protein expression was induced by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside. Cells were incubated at 25 °C and harvested after 16 h by centrifuging at 5000 x *g* and 4 °C for 15 min. The cell paste was resuspended in Buffer C [20 mM Tris-HCl pH 8.0, 200 mM NaCl, 5% (v/v) glycerol, 10 mM imidazole] and disrupted by sonication. The resulting suspension was centrifuged twice at 30000 x *g* and 4 °C for 15 min and the soluble fraction (crude extract) was loaded onto a 5 ml HisTrap column (GE Life Sciences, Piscataway, NJ) containing Ni^{2+} and previously equilibrated with Buffer C. Elusion of the retained proteins was achieved with a linear imidazole gradient (20 column volumes, 10-300 mM). Fractions containing SucSase activity were pooled, concentrated to 2 ml, and loaded onto a 16/60 Superdex 200 column (GE Life Sciences) previously equilibrated with 50 mM HEPES-NaOH pH 8.0 and 300 mM NaCl. Fractions containing enzyme activity were pooled, concentrated, supplemented with 5% (v/v) glycerol, and stored at -80 °C until use. Under these conditions the enzyme remained stable and fully active for at least 3 months.

Protein assay and detection: Protein concentration was determined by measuring the protein absorbance at 280 nm using a NanoDrop 1000 (Thermo Fisher Scientific) and an extinction coefficient of 1.153 ml mg^{-1} cm^{-1}, determined from the amino acid sequence using the ProtParam server (http://web.expasy.org/protparam/). Denaturing protein electrophoresis was performed as described by Laemmli [100].

Enzyme assays: Activity assays were performed as previously described [77], with minor modifications. In the direction of Suc synthesis, the reaction medium contained 50 mM HEPPS pH 8.0, 10 mM $MgCl_2$, 5 mM UDP-Glc, 500 mM Fru, 0.3 mM

phosphoenolpyruvate, 0.3 mM NADH, 1 U pyruvate kinase, 1 U lactate dehydrogenase, 0.2 mg ml^{-1} BSA, and enzyme in an appropriate dilution in a final volume of 50 µl. Alternatively, activity was measured with 1 mM ADP-Glc and 20 mM Fru. NADH oxidation was followed by measuring the absorbance at 340 nm in a Multiskan Ascent microplate reader (Thermo Fisher Scientific) at 37 °C. One unit of enzyme activity (U) is defined as the amount of protein necessary to produce 1 µmol of product in 1 min under the specified conditions.

Kinetic characterization: Data of initial velocity (v) versus substrate concentration (S) were plotted and fitted to a modified Hill equation: $v = V_{max}\ S^{nH} / (S_{0.5}{}^{nH} + S^{nH})$, where $S_{0.5}$ is the concentration of substrate necessary to obtain 50% of the maximal velocity (V_{max}) and n_H is the Hill coefficient. Fitting was performed by a non-linear least-squares algorithm provided by the software Origin 7.0 (OriginLab Corporation). Kinetic parameters were obtained using the averages of two independent datasets that were reproducible within errors of ± 10%.

Phylogenetic analysis: We searched for protein sequences using the term "sucrose synthase" and the RefSeq filter in the National Center for Biotechnology Information (NCBI) database. Sequences were analyzed with the program BioEdit 7.0.5.3 [101], manually curated, and aligned using the ClustalW server (http://www.genome.jp/tools/clustalw/). Tree reconstruction was performed using the Neighbor-Joining algorithm with a bootstrap of 1000 in the program SeaView 4.4.0 [102]. The tree figure was prepared using the FigTree 1.4.0 software (http://tree.bio.ed.ac.uk/software/figtree/).

Crystallization and data collection: After the initial crystallization screen and optimization, the recombinant protein was crystallized via the hanging drop method. The hanging drops were prepared with 1 µl of 15 mg ml^{-1} *Neu*SucSase protein solution and 1 µl of the reservoir solution, containing 5% Tacsimate pH 5.0, 5% (w/v) PEG 3350, and 0.1 M sodium citrate pH 5.6. The hanging drops were kept at 20 °C for crystallization. Crystals appeared in 3 days and were allowed to continue growing at 20 °C for 4 more days until they reached their maximum sizes. Crystals with good morphology and large sizes were transferred to a cryo-condition, which contained 25% glycerol in addition to the components of the reservoir solution, before being frozen in liquid nitrogen.
X-ray diffraction data sets were collected at the SBC19-ID beamline at the Advanced Photon Source (Argonne National Laboratory, Chicago, IL). The wavelength used in the monochromatic data collection was 1.008 Å. All the collected data sets were indexed and integrated using iMosflm and scaled with Scala in the CCP4 program suite (Collaborative Computational Project Number 4) [103]. The resolution cut-off was decided by applying all following criteria on the highest resolution bin: i) I/Sigma > 2, ii) completeness > 90%, and iii) $R_{merge} < 1.0$.

Phasing, model building, and refinement: Molecular replacement was carried out using the program Phaser [61] from the CCP4 program suite. The starting search model in molecular replacement was modified from the known *A. thaliana* SucSase structural model (PDB ID: 3S29) [84]. We truncated the GT-B(D) domain (cyan domain in Figure 36B) and used the rest of the molecule as the search model for molecular replacement in Phaser. Once a solution was obtained, model building was conducted in COOT [45]. The

GT-B(D) domain was built according to the electron density maps. Rigid body refinement and restrained refinement were conducted in refmac5 [104]. In order to remove model bias and achieve the best refinement results possible, simulated annealing refinement and ordered solvent identification were conducted using PHENIX.refine [44]. Final model and the structure factor have been deposited in the RCSB Protein Data Bank with the accession code 4RBN.

Homology modeling: A model of the monomeric closed form of *Neu*SucSase (residues 16 to 788) was constructed with the program Modeller 9.11 (http://salilab.org/modeller/) [105]. As template we used the atomic coordinates of the *A. thaliana* sucrose synthase (3S27) with the ligands UDP and fructose [84]. Before the modeling process, sequence alignment was performed manually to match functionally conserved residues and secondary structures. An identity of 52.2% ensured a high confidence alignment since we only had to introduce four one-residue indels. The accuracy of the models was assessed with the Verify3D Structure Evaluation Server (http://nihserver.mbi.ucla.edu/Verify_3D/) [106].

Difference distance matrix map: We used and *ad hoc* program written in C applying previously developed concepts to detect domain motion and identify regions that move closer upon conformational changes [107]. Distances were calculated between all pair of C_α of one reference structure (open), and a second pairwise distance matrix was calculated for the target (closed) structure. Afterwards, the target matrix was subtracted from the reference matrix to calculate the Δdistance plot (https://github.com/ballicoragroup/didimama).

Hinge analysis: In order to detect possible local conformations or hinges, we performed an analysis with the ad hoc program "hingescan" (https://github.com/ballicoragroup/hingescan). We compared the open crystal structure of *Neu*SucSas*e* with a *Neu*SucSas*e* homology model based on the template provided by the close structure of the *Ath*SucSase. To detect if there is a significant local conformational change around a given residue ("hinge"), we extracted the coordinates of a given number (n) of C_α before the putative hinge and the same given number (n) of residues after (window size = 2n+1). This was done for both the open and closed forms and obtained two fragments to compare. After optimal rigid body superposition of only these two set of coordinates, an average distance was calculated (RMSD). This RMSD calculated in these conditions was called the "hinge score". When this score is at a peak, the "flanking" n number of C_α at both sides display a maximum change between the two structures. For that reason, a hinge is detected. The bigger the window, the bigger the domain movement is detected surrounding the hinge. To identify hinges that link small and bigger domains, different window sizes can be scanned.

CHAPTER SIX

CONCLUSIONS

We analyzed the sequences of SucSases from divergent organisms, including proteobacteria, cyanobacteria, and plants. The obtained phylogenetic tree shows that sequences from proteobacteria (including the one from *N. europaea*) are clearly separated from those present in cyanobacteria and plants. To the best of our knowledge, SucSases from proteobacteria have never been characterized. In this work, we expressed, purified the recombinant *Neu*SucSase, and solved its crystal structure. In addition, our analysis provided evidence for a proposed "open" conformation for SucSases. Based on the comparison with a previously published "closed" *Ath*SucSase structure [84], a "hinge-latch" combination was identified as critical features responsible for the open-close enzyme actions.

We also mutated three highly conserved amino acids proposed to be critical for catalysis. These mutations severely reduced the activity of *Neu*SucSase; therefore, we conclude that the triad composed of residues R567, K572, and E663 (numbers according to *Neu*SucSase) plays a key role for catalysis not only in SucSases, but probably also in all the retaining GT-B glycosyltransferases [84, 94-97].

With both structural and kinetic results we propose that *Neu*SucSase has a substrate preference in favor of ADP/ADP-Glc over UDP/UDP-Glc. This behavior is

similar to the one observed for *T. elongatus* SucSase [77]. It was proposed that the difference in specificity towards the nucleotide substrate in the cyanobacterial enzyme could be due to changes in two specific residues, R647 and S653 (Q648 and N654 in *Ath*SucSase). Interestingly, we found that *Neu*SucSase presents similar evolutionary variances to those observed in *T. elongatus* SucSase: R636 (R647) and A642 (S653). Residues Q648 and N654 from *Ath*SucSase are strictly conserved in plants and the moss *P. patens*, whereas they are more variable in cyanobacteria and proteobacteria. Based on this finding and taking into account kinetic and structural data obtained for SucSases from *T. elongatus* [77], *Ath*SucSase [84] and *Neu*SucSase (this work), we hypothesize that these residues may contribute to determining the specificity towards the nucleotide substrate.

The evolutionary origin of enzymes from Suc metabolism in proteobacteria has been previously discussed [65, 66, 69]. The evolution of SucSases in cyanobacteria, proteobacteria, and plants is not yet fully understood, but most likely it involved horizontal gene transfers. On one hand, *Neu*SucSase is closer to plant enzymes in the phylogenetic tree (Figure 35), but on the other hand, the specificity for nucleotides is similar to several cyanobacterial enzymes examined [69, 77]. It is possible that *Neu*SucSase evolved from a protein already present in the common ancestor of proteobacteria and cyanobacteria [71].

REFERENCE LIST

[1] Davidson, D. J., and Porteous, D. J. (1998) Genetics and pulmonary medicine. 1. The genetics of cystic fibrosis lung disease, *Thorax 53*, 389-397.

[2] Rojo, F. (1999) Repression of transcription initiation in bacteria, *Journal of bacteriology 181*, 2987-2991.

[3] Lewis, P. J., Doherty, G. P., and Clarke, J. (2008) Transcription factor dynamics, *Microbiology 154*, 1837-1844.

[4] Grose, J. H., Bergthorsson, U., and Roth, J. R. (2005) Regulation of NAD synthesis by the trifunctional NadR protein of Salmonella enterica, *Journal of bacteriology 187*, 2774-2782.

[5] Browning, D. F., and Busby, S. J. (2004) The regulation of bacterial transcription initiation, *Nature reviews. Microbiology 2*, 57-65.

[6] Haydon, D. J., and Guest, J. R. (1991) A new family of bacterial regulatory proteins, *FEMS microbiology letters 63*, 291-295.

[7] Yoshida, K. I., Fujita, Y., and Ehrlich, S. D. (2000) An operon for a putative ATP-binding cassette transport system involved in acetoin utilization of Bacillus subtilis, *Journal of bacteriology 182*, 5454-5461.

[8] Klemm, J. D., Schreiber, S. L., and Crabtree, G. R. (1998) Dimerization as a regulatory mechanism in signal transduction, *Annual review of immunology 16*, 569-592.

[9] Littlefield, O., and Nelson, H. C. (1999) A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocrystal, *Nature structural biology 6*, 464-470.

[10] Vindal, V., Suma, K., and Ranjan, A. (2007) GntR family of regulators in Mycobacterium smegmatis: a sequence and structure based characterization, *BMC genomics 8*, 289.

[11] Park, Y. W., Yeo, H. K., and Lee, J. Y. (2012) Crystallization and preliminary X-ray diffraction analysis of a fatty-acid metabolism regulatory protein, FadR, from Bacillus halodurans, *Acta crystallographica. Section F, Structural biology and crystallization communications 68*, 975-977

[12] Allison, S. L., and Phillips, A. T. (1990) Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of Pseudomonas putida, *Journal of bacteriology 172*, 5470-5476.

[13] Quail, M. A., Dempsey, C. E., and Guest, J. R. (1994) Identification of a fatty acyl responsive regulator (FarR) in Escherichia coli, *FEBS letters 356*, 183-187.

[14] Chen, C. M., Ye, Q. Z., Zhu, Z. M., Wanner, B. L., and Walsh, C. T. (1990) Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psiD) genes involved in alkylphosphonate uptake and C-P lyase activity in Escherichia coli B, *The Journal of biological chemistry 265*, 4461-4471.

[15] Resch, M., Schiltz, E., Titgemeyer, F., and Muller, Y. A. (2010) Insight into the induction mechanism of the GntR/HutC bacterial transcription regulator YvoA, *Nucleic acids research 38*, 2485-2497.

[16] Rigali, S., Derouaux, A., Giannotta, F., and Dusart, J. (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies, *The Journal of biological chemistry 277*, 12507-12515.

[17] Gao, Y. G., Yao, M., Itou, H., Zhou, Y., and Tanaka, I. (2007) The structures of transcription factor CGL2947 from Corynebacterium glutamicum in two crystal forms: a novel homodimer assembling and the implication for effector-binding mode, *Protein science : a publication of the Protein Society 16*, 1878-1886.

[18] Hoskisson, P. A., Rigali, S., Fowler, K., Findlay, K. C., and Buttner, M. J. (2006) DevA, a GntR-like transcriptional regulator required for development in Streptomyces coelicolor, *Journal of bacteriology 188*, 5014-5023.

[19] Franco, I. S., Mota, L. J., Soares, C. M., and de Sa-Nogueira, I. (2006) Functional domains of the Bacillus subtilis transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding, *Journal of bacteriology 188*, 3024-3036.

[20] Jain, D., and Nair, D. T. (2013) Spacing between core recognition motifs determines relative orientation of AraR monomers on bipartite operators, *Nucleic acids research 41*, 639-647.

[21] Kim, J. S., Kim, B. H., Jang, J. I., Eom, J. S., Kim, H. G., Bang, I. S., and Park, Y. K. (2014) Functional insight from the tetratricopeptide repeat-like motifs of the type III secretion chaperone SicA in Salmonella enterica serovar Typhimurium, *FEMS microbiology letters 350*, 146-153.

[22] Lee, M. H., Scherer, M., Rigali, S., and Golden, J. W. (2003) PlmA, a new member of the GntR family, has plasmid maintenance functions in Anabaena sp. strain PCC 7120, *Journal of bacteriology 185*, 4315-4325.

[23] Wiethaus, J., Schubert, B., Pfander, Y., Narberhaus, F., and Masepohl, B. (2008) The GntR-like regulator TauR activates expression of taurine utilization genes in Rhodobacter capsulatus, *Journal of bacteriology 190*, 487-493.

[24] Belitsky, B. R. (2004) Bacillus subtilis GabR, a protein with DNA-binding and aminotransferase domains, is a PLP-dependent transcriptional regulator, *Journal of molecular biology 340*, 655-664.

[25] Belitsky, B. R., and Sonenshein, A. L. (2002) GabR, a member of a novel protein family, regulates the utilization of gamma-aminobutyrate in Bacillus subtilis, *Molecular microbiology 45*, 569-583.

[26] Bramucci, E., Milano, T., and Pascarella, S. (2011) Genomic distribution and heterogeneity of MocR-like transcriptional factors containing a domain belonging to the superfamily of the pyridoxal-5'-phosphate dependent enzymes of fold type I, *Biochemical and biophysical research communications 415*, 88-93.

[27] Kirsch, J. F., Eichele, G., Ford, G. C., Vincent, M. G., Jansonius, J. N., Gehring, H., and Christen, P. (1984) Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure, *Journal of molecular biology 174*, 497-525.

[28] Schneider, G., Kack, H., and Lindqvist, Y. (2000) The manifold of vitamin B6 dependent enzymes, *Structure 8*, R1-6.

[29] Ouchi, T., Tomita, T., Miyagawa, T., Kuzuyama, T., and Nishiyama, M. (2009) Dual roles of a conserved pair, Arg23 and Ser20, in recognition of multiple substrates in alpha-aminoadipate aminotransferase from Thermus thermophilus, *Biochemical and biophysical research communications 388*, 21-27.

[30] Gajiwala, K. S., and Burley, S. K. (2000) Winged helix proteins, *Current opinion in structural biology 10*, 110-116.

[31] Eliot, A. C., and Kirsch, J. F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations, *Annual review of biochemistry 73*, 383-415.

[32] Smith, D. L., Almo, S. C., Toney, M. D., and Ringe, D. (1989) 2.8-A-resolution crystal structure of an active-site mutant of aspartate aminotransferase from Escherichia coli, *Biochemistry 28*, 8161-8167.

[33] Storici, P., Qiu, J., Schirmer, T., and Silverman, R. B. (2004) Mechanistic crystallography. Mechanism of inactivation of gamma-aminobutyric acid aminotransferase by (1R,3S,4S)-3-amino-4-fluorocyclopentane-1-carboxylic acid as elucidated by crystallography, *Biochemistry 43*, 14057-14063.

[34] Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C., and Bryant, S. H. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins, *Nucleic acids research 39*, D225-229.

[35] van Aalten, D. M., DiRusso, C. C., and Knudsen, J. (2001) The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR, *The EMBO journal 20*, 2041-2050.

[36] Raman, N., Black, P. N., and DiRusso, C. C. (1997) Characterization of the fatty acid-responsive transcription factor FadR. Biochemical and genetic analyses of the native conformation and functional domains, *The Journal of biological chemistry 272*, 30645-30650.

[37] Storici, P., De Biase, D., Bossa, F., Bruno, S., Mozzarelli, A., Peneff, C., Silverman, R. B., and Schirmer, T. (2004) Structures of gamma-aminobutyric acid (GABA) aminotransferase, a pyridoxal 5'-phosphate, and [2Fe-2S] cluster-containing enzyme, complexed with gamma-ethynyl-GABA and with the antiepilepsy drug vigabatrin, *The Journal of biological chemistry 279*, 363-373.

[38] Tomchick, D. R., Turner, R. J., Switzer, R. L., and Smith, J. L. (1998) Adaptation of an enzyme to regulatory function: structure of Bacillus subtilis PyrR, a pyr RNA-binding attenuation protein and uracil phosphoribosyltransferase, *Structure 6*, 337-350.

[39] Sinha, S. C., Krahn, J., Shin, B. S., Tomchick, D. R., Zalkin, H., and Smith, J. L. (2003) The purine repressor of Bacillus subtilis: a novel combination of domains adapted for transcription regulation, *Journal of bacteriology 185*, 4087-4098.

[40] Jain, D., Kim, Y., Maxwell, K. L., Beasley, S., Zhang, R., Gussin, G. N., Edwards, A. M., and Darst, S. A. (2005) Crystal structure of bacteriophage lambda cII and its DNA complex, *Molecular cell 19*, 259-269.

[41] Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol 276*, 307-326.

[42] Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography, *Acta crystallographica. Section D, Biological crystallography 50*, 760-763.

[43] Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic acids research 38*, W545-549.

[44] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr. 66*, 213-221.

[45] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126-2132.

[46] Edayathumangalam, R., Wu, R., Garcia, R., Wang, Y., Wang, W., Kreinbring, C. A., Bach, A., Liao, J., Stone, T. A., Terwilliger, T. C., Hoang, Q. Q., Belitsky, B. R., Petsko, G. A., Ringe, D., and Liu, D. (2013) Crystal structure of Bacillus subtilis GabR, an autorepressor and transcriptional activator of gabT, *Proceedings of the National Academy of Sciences of the United States of America 110*, 17820-17825.

[47] Jochmann, N., Gotker, S., and Tauch, A. (2011) Positive transcriptional control of the pyridoxal phosphate biosynthesis genes pdxST by the MocR-type regulator PdxR of Corynebacterium glutamicum ATCC 13032, *Microbiology 157*, 77-88.

[48] Belitsky, B. R. (2014) Role of PdxR in the activation of vitamin B6 biosynthesis in Listeria monocytogenes, *Molecular microbiology 92*, 1113-1128.

[49] El Qaidi, S., Yang, J., Zhang, J. R., Metzger, D. W., and Bai, G. (2013) The vitamin B(6) biosynthesis pathway in Streptococcus pneumoniae is controlled by pyridoxal 5'-phosphate and the transcription factor PdxR and has an impact on ear infection, *Journal of bacteriology 195*, 2187-2196.

[50] Liu, W., Peterson, P. E., Langston, J. A., Jin, X., Zhou, X., Fisher, A. J., and Toney, M. D. (2005) Kinetic and crystallographic analysis of active site mutants of Escherichia coli gamma-aminobutyrate aminotransferase, *Biochemistry 44*, 2982-2992.

[51] Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure

determination, *Acta crystallographica. Section D, Biological crystallography 58*, 1948-1954.

[52] Liu, W., Peterson, P. E., Carter, R. J., Zhou, X., Langston, J. A., Fisher, A. J., and Toney, M. D. (2004) Crystal structures of unbound and aminooxyacetate-bound Escherichia coli gamma-aminobutyrate aminotransferase, *Biochemistry 43*, 10896-10905.

[53] Silverman, R. B., and Levy, M. A. (1981) Mechanism of inactivation of gamma-aminobutyric acid-alpha-ketoglutaric acid aminotransferase by 4-amino-5-halopentanoic acids, *Biochemistry 20*, 1197-1203.

[54] Liu, D., Pozharski, E., Fu, M., Silverman, R. B., and Ringe, D. (2010) Mechanism of inactivation of Escherichia coli aspartate aminotransferase by (S)-4-amino-4,5-dihydro-2-furancarboxylic acid, *Biochemistry 49*, 10507-10515.

[55] Liu, D., Pozharski, E., Lepore, B. W., Fu, M., Silverman, R. B., Petsko, G. A., and Ringe, D. (2007) Inactivation of Escherichia coli L-aspartate aminotransferase by (S)-4-amino-4,5-dihydro-2-thiophenecarboxylic acid reveals "a tale of two mechanisms", *Biochemistry 46*, 10517-10527.

[56] Lepore, B. W., Liu, D., Peng, Y., Fu, M., Yasuda, C., Manning, J. M., Silverman, R. B., and Ringe, D. (2010) Chiral discrimination among aminotransferases: inactivation by 4-amino-4,5-dihydrothiophenecarboxylic acid, *Biochemistry 49*, 3138-3147.

[57] Schnackerz, K. D., Tai, C. H., Simmons, J. W., 3rd, Jacobson, T. M., Rao, G. S., and Cook, P. F. (1995) Identification and spectral characterization of the external aldimine of the O-acetylserine sulfhydrylase reaction, *Biochemistry 34*, 12152-12160.

[58] Griswold, W. R., Castro, J. N., Fisher, A. J., and Toney, M. D. (2012) Ground-state electronic destabilization via hyperconjugation in aspartate aminotransferase, *Journal of the American Chemical Society 134*, 8436-8438.

[59] Griswold, W. R., Fisher, A. J., and Toney, M. D. (2011) Crystal structures of aspartate aminotransferase reconstituted with 1-deazapyridoxal 5'-phosphate: internal aldimine and stable L-aspartate external aldimine, *Biochemistry 50*, 5918-5924.

[60] Onuffer, J. J., and Kirsch, J. F. (1995) Redesign of the substrate specificity of Escherichia coli aspartate aminotransferase to that of Escherichia coli tyrosine aminotransferase by homology modeling and site-directed mutagenesis, *Protein science : a publication of the Protein Society 4*, 1750-1757.

[61] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of applied crystallography 40*, 658-674.

[62] Koch, K. (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development, *Current opinion in plant biology 7*, 235-246.

[63] Lunn, J. E. (2008) Sucrose Metabolism, In *Encyclopedia of Life Sciences*, John Wiley & Sons.

[64] Winter, H., and Huber, S. C. (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes, *Critical reviews in biochemistry and molecular biology 35*, 253-289.

[65] MacRae, E., and Lunn, J. (2012) Photosynthetic Sucrose Biosynthesis: An Evolutionary Perspective, In *Photosynthesis* (Eaton-Rye, J. J., Tripathy, B. C., and Sharkey, T. D., Eds.), pp 675-702, Springer Netherlands.

[66] Salerno, G. L., and Curatti, L. (2003) Origin of sucrose metabolism in higher plants: when, how and why?, *Trends in plant science 8*, 63-69.

[67] Cumino, A. C., Marcozzi, C., Barreiro, R., and Salerno, G. L. (2007) Carbon cycling in *Anabaena* sp. PCC 7120. Sucrose synthesis in the heterocysts and possible role in nitrogen fixation, *Plant physiology 143*, 1385-1397.

[68] Curatti, L., Giarrocco, L., and Salerno, G. L. (2006) Sucrose synthase and RuBisCo expression is similarly regulated by the nitrogen source in the nitrogen-fixing cyanobacterium *Anabaena* sp., *Planta 223*, 891-900.

[69] Kolman, M. A., Torres, L. L., Martin, M. L., and Salerno, G. L. (2012) Sucrose synthase in unicellular cyanobacteria and its relationship with salt and hypoxic stress, *Planta 235*, 955-964.

[70] Yamanaka, T. (2008) *Chemolithoautotrophic Bacteria: Biochemistry and Environmental Biology*, Springer, Japan.

[71] Teske, A., Alm, E., Regan, J. M., Toze, S., Rittmann, B. E., and Stahl, D. A. (1994) Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria, *J. Bacteriol. 176*, 6623-6630.

[72] Vannelli, T., Logan, M., Arciero, D. M., and Hooper, A. B. (1990) Degradation of halogenated aliphatic compounds by the ammonia- oxidizing bacterium *Nitrosomonas europaea*, *Applied and environmental microbiology 56*, 1169-1171.

[73] Chain, P., Kurtz, S., Ohlebusch, E., and Slezak, T. (2003) An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges, *Briefings in bioinformatics 4*, 105-123.

[74] Hooper, A. B. (1969) Biochemical basis of obligate autotrophy in *Nitrosomonas europaea*, *J. Bacteriol. 97*, 776-779.

[75] Baroja-Fernandez, E., Munoz, F. J., Saikusa, T., Rodriguez-Lopez, M., Akazawa, T., and Pozueta-Romero, J. (2003) Sucrose synthase catalyzes the de novo production of ADPglucose linked to starch biosynthesis in heterotrophic tissues of plants, *Plant & cell physiology 44*, 500-509.

[76] Delmer, D. P. (1972) The purification and properties of sucrose synthetase from etiolated *Phaseolus aureus* seedlings, *J. Biol. Chem. 247*, 3822-3828.

[77] Figueroa, C. M., Asencion Diez, M. D., Kuhn, M. L., McEwen, S., Salerno, G. L., Iglesias, A. A., and Ballicora, M. A. (2013) The unique nucleotide specificity of the sucrose synthase from Thermosynechococcus elongatus, *FEBS letters 587*, 165-169.

[78] Grimes, W. J., Jones, B. L., and Albersheim, P. (1970) Sucrose synthetase from *Phaseolus aureus* seedlings, *J. Biol. Chem. 245*, 188-197.

[79] Morell, M., and Copeland, L. (1985) Sucrose synthase of soybean nodules, *Plant physiology 78*, 149-154.

[80] Porchia, A. C., Curatti, L., and Salerno, G. L. (1999) Sucrose metabolism in cyanobacteria: sucrose synthase from *Anabaena* sp. strain PCC 7119 is remarkably different from the plant enzymes with respect to substrate affinity and amino-terminal sequence, *Planta 210*, 34-40.

[81] Ross, H. A., and Davies, H. V. (1992) Purification and Characterization of Sucrose Synthase from the Cotyledons of *Vicia faba* L., *Plant physiology 100*, 1008-1013.

[82] Tanase, K., and Yamaki, S. (2000) Purification and characterization of two sucrose synthase isoforms from Japanese pear fruit, *Plant & cell physiology 41*, 408-414.

[83] Curatti, L., Giarrocco, L. E., Cumino, A. C., and Salerno, G. L. (2008) Sucrose synthase is involved in the conversion of sucrose to polysaccharides in filamentous nitrogen-fixing cyanobacteria, *Planta 228*, 617-625.

[84] Zheng, Y., Anderson, S., Zhang, Y., and Garavito, R. M. (2011) The structure of sucrose synthase-1 from *Arabidopsis thaliana* and its functional implications, *J. Biol. Chem. 286*, 36108-36118.

[85] Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms, *Annu. Rev. Biochem. 77*, 521-555.

[86] Sheng, F., Jia, X., Yep, A., Preiss, J., and Geiger, J. H. (2009) The crystal structures of the open and catalytically competent closed conformation of *Escherichia coli* glycogen synthase, *J. Biol. Chem. 284*, 17796-17807.

[87] Hardin, S. C., Winter, H., and Huber, S. C. (2004) Phosphorylation of the amino terminus of maize sucrose synthase in relation to membrane association and enzyme activity, *Plant physiology 134*, 1427-1438.

[88] Klotz, K. L., Finger, F. L., and Shelver, W. L. (2003) Characterization of two sucrose synthase isoforms in sugarbeet root, *Plant Physiol. Biochem. 41*, 107-115.

[89] Schäfer, W. E., Rohwer, J. M., and Botha, F. C. (2004) A kinetic study of sugarcane sucrose synthase, *Eur. J. Biochem. 271*, 3971-3977.

[90] Wolosiuk, R. A., and Pontis, H. G. (1974) Studies on sucrose synthetase. Kinetic mechanism, *Archives of biochemistry and biophysics 165*, 140-145.

[91] Gibson, R. P., Tarling, C. A., Roberts, S., Withers, S. G., and Davies, G. J. (2004) The donor subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 A resolution, *The Journal of biological chemistry 279*, 1950-1955.

[92] Yep, A., Ballicora, M. A., and Preiss, J. (2006) The ADP-glucose binding site of the Escherichia coli glycogen synthase, *Archives of biochemistry and biophysics 453*, 188-196.

[93] Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A., and Alzari, P. M. (2004) Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation, *The EMBO journal 23*, 3196-3205.

[94] Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R., and Davies, G. J. (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA, *Chemistry & biology 9*, 1337-1346.

[95] Schinzel, R., and Drueckes, P. (1991) The phosphate recognition site of Escherichia coli maltodextrin phosphorylase, *FEBS letters 286*, 125-128.

[96] Schinzel, R., and Palm, D. (1990) Escherichia coli maltodextrin phosphorylase: contribution of active site residues glutamate-637 and tyrosine-538 to the phosphorolytic cleavage of alpha-glucans, *Biochemistry 29*, 9956-9962.

[97] Yep, A., Ballicora, M. A., and Preiss, J. (2004) The active site of the *Escherichia coli* glycogen synthase is similar to the active site of retaining GT-B glycosyltransferases, *Biochem. Biophys. Res. Commun. 316*, 960-966.

[98] Kuhn, M. L., Falaschetti, C. A., and Ballicora, M. A. (2009) Ostreococcus tauri ADP-glucose pyrophosphorylase reveals alternative paths for the evolution of subunit roles, *The Journal of biological chemistry 284*, 34092-34102.

[99] Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A laboratory Manual*, Vol. 2, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[100] Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680-685.

[101] Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucl. Acids Symp. Ser. 41*, 95-98.

[102] Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building, *Molecular biology and evolution 27*, 221-224.

[103] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta. Crystallogr. D Biol. Crystallogr. 67*, 235-242.

[104] Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D Biol. Crystallogr. 53*, 240-255.

[105] Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *Journal of molecular biology 234*, 779-815.

[106] Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles, *Nature 356*, 83-85.

[107] Richards, F. M., and Kundrot, C. E. (1988) Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure, *Proteins 3*, 71-84.

VITA

Rui Wu was born in Linfen, Shanxi, China. She was 23 when moved to Chicago, Illinois, where she attended graduate school. Between 2006 and 2010 she attended Sichuan University, Chengdu, China, where she earned a Bachelor of Science in